

# A human-specific regulatory mechanism revealed in a pre-implantation model

<https://doi.org/10.1038/s41586-025-09571-1>

Received: 28 March 2024

Accepted: 22 August 2025

Published online: 01 October 2025

Open access

 Check for updates

Raquel Fueyo<sup>1</sup>, Sicong Wang<sup>2</sup>, Olivia J. Crocker<sup>3</sup>, Tomek Swigut<sup>1</sup>, Hiromitsu Nakauchi<sup>2,4</sup> & Joanna Wysocka<sup>1,2,5,6</sup>✉

Stem cell-based human embryo models offer a unique opportunity for functional studies of the human-specific features of development. Here we genetically and epigenetically manipulate human blastoids, a 3D embryo model of the blastocyst<sup>1</sup>, to investigate the functional effect of HERVK LTR5Hs, a hominoid-specific endogenous retrovirus, on pre-implantation development. We uncover a pervasive *cis*-regulatory contribution of LTR5Hs elements to the hominoid-specific diversification of the epiblast transcriptome in blastoids. Many of the LTR5Hs genomic insertions in the human genome are unique to our own species. We show that at least one such human-specific LTR5Hs element is essential for the blastoid-forming potential via enhancing expression of the primate-specific *ZNF729* gene, encoding a KRAB zinc-finger protein. *ZNF729* binds to GC-rich sequences, abundant at gene promoters associated with basic cellular functions, such as cell proliferation and metabolism. Despite mediating recruitment of TRIM28, at many of these promoters *ZNF729* acts as a transcriptional activator. Together, our results illustrate how recently emerged transposable elements and genes can confer developmentally essential functions in humans.

Endogenous retroviruses (ERVs), also called long terminal repeat (LTR) retrotransposons, comprise approximately 8.9% of the human genome<sup>2</sup>. ERVs are remnants of ancient retroviral infections of the germline that integrated into the genome, propagated vertically and ultimately became fixed in the population<sup>3</sup>. To successfully endogenize, the LTRs of ERVs must have entered the host genome able to engage the transcriptional machinery in germ cells or pre-implantation embryo cells. This, and the reduced DNA methylation, can potentially account for the widespread regulatory co-option of ERVs in mammalian pre-implantation development<sup>4</sup>, a period that spans the time from fertilization to the attachment of the blastocyst to the uterus. For example, during mouse pre-implantation, many LTRs function as stage-specific promoters<sup>5–8</sup>. In humans, ERVs of the HERVK (HML-2) family—specifically those of the LTR5Hs subtype—are transcriptionally activated in the embryo around the eight-cell stage and stay active in the blastocyst<sup>9</sup> (Extended Data Fig. 1a). HERVK LTR5Hs is also active in human teratocarcinoma and human naive pluripotent stem cells (hnPSCs) where it exerts enhancer function<sup>10,11</sup>. HERVK LTR5Hs is the evolutionarily most recent ERV in humans. It first invaded the genome after the split of hominoids (apes) from Old World monkeys, and it remained active following the split of humans and chimpanzees<sup>12</sup>. As a result, the approximately 700 LTR5Hs insertions in the human genome are unique to hominoids, with a subset being specific to humans<sup>12,13</sup>. Nonetheless, the functional effect of HERVK LTR5Hs on pre-implantation development remains poorly understood.

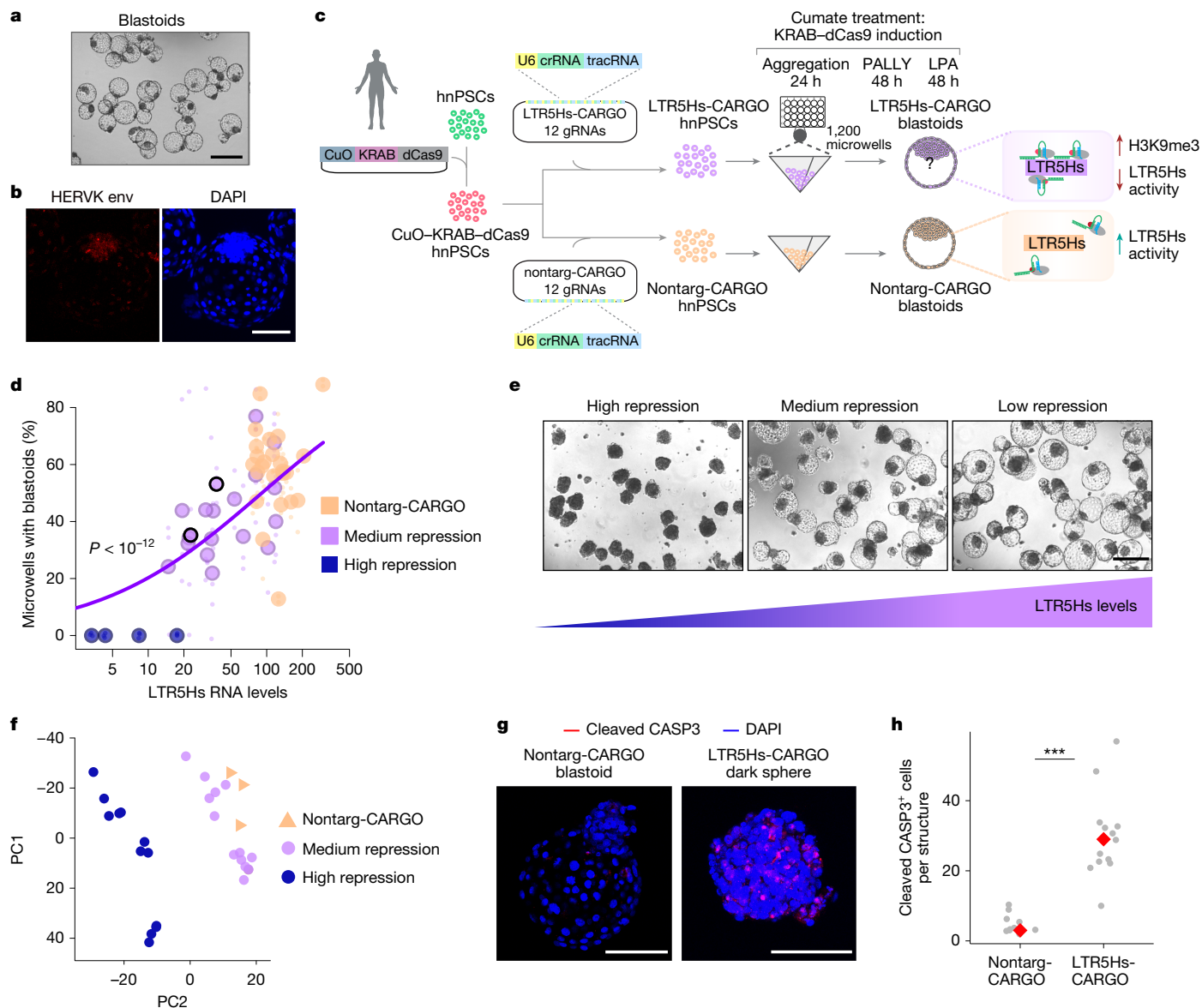
Although the general principles of early development are conserved across mammals, many aspects have diverged between species<sup>14,15</sup>.

Ethical and practical considerations limit functional studies in human embryos<sup>16</sup>. However, groundbreaking work established stem cell-based 3D embryo models named blastoids, which recapitulate the morphology and lineage specification of the human blastocyst (reviewed in ref. 1). Although not without limitations, human blastoids offer unprecedented opportunities to study species-specific features of human pre-implantation. Here we perturbed the function of HERVK LTR5Hs in human blastoids and reveal its dose-dependent effect on blastoid formation and gene regulation. We further uncovered a human-specific LTR5Hs insertion that enhances expression of the gene encoding the zinc-finger transcription factor *ZNF729*, promotes proliferation of hnPSCs and is essential for endowing hnPSCs with blastoid-forming potential. We demonstrate that *ZNF729* binds to and regulates GC-rich promoters of genes involved in fundamental cellular functions. Together, our work reveals an evolutionary novel mechanism regulating conserved cellular processes and paves the way for systematic interrogation of transposable element function using *in vitro* human embryo models.

## HERVK LTR5Hs impacts blastoid formation

To probe the phenotypic effect of HERVK LTR5Hs activity on human pre-implantation, we turned to a human blastoid model<sup>17</sup> (Fig. 1a). Blastoids contain analogues to the blastocyst lineages: the epiblast that gives rise to the embryo proper, the trophectoderm that generates the placenta, and the hypoblast that develops into the yolk sac<sup>14</sup>. Analysis of the HERVK expression pattern in human embryos and blastoids using published single-cell RNA sequencing (scRNA-seq)

<sup>1</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA. <sup>2</sup>Institute of Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>3</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. <sup>4</sup>Stem Cell Therapy Laboratory, Advanced Research Institute, Tokyo Medical and Dental University, Tokyo, Japan. <sup>5</sup>Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. <sup>6</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA. ✉e-mail: wysocka@stanford.edu



**Fig. 1 | LTR5Hs activity contributes to the blastoid formation potential of hnPSCs.** **a**, Representative bright-field image of wild-type blastoids ( $n = 3$  biological replicates). Scale bar, 400  $\mu\text{m}$ . **b**, Representative confocal images of a wild-type blastoid immunostained with HERVK envelope (env) antibodies and DAPI ( $n = 4$  biological replicates). Scale bar, 100  $\mu\text{m}$ . **c**, Schematic of nontarg-CARGO and LTR5Hs-CARGO hnPSCs and the generation of blastoids. The cartoon summarizes how LTR5Hs-CARGO promotes H3K9me3 deposition and LTR5Hs repression (right). crRNA, CRISPR RNA; LPA, media containing 1-oleoyl lysophosphatidic acid sodium salt; PALLY, media containing PDO325901, A83-01, leukemia inhibitory factor, LPA and Y-27632; tracRNA, *trans*-activating crRNA. The silhouette of the human was created in BioRender. Fueyo, R. (2025) <https://BioRender.com/tmgd0pn>. **d**, Blastoid-forming potential of hnPSC depends on LTR5Hs activity ( $\beta$ -regression, two-sided log-likelihood test =  $2.39 \times 10^{-13}$ ). Blastoid formation was assessed in 23 LTR5Hs-CARGO ( $n = 3$  biological replicates) and 24 nontarg-CARGO ( $n = 2$  biological replicates) clonal cell lines. Plotted is the proportion of blastoids per well versus Taqman LTR5Hs

RNA levels normalized to *RPL13A* RNA levels. The purple line represents the LTR5Hs-CARGO regression line. The purple circles with black outlines indicate clones used for scRNA-seq experiments. **e**, Bright-field images of structures collected upon LTR5Hs high repression (left; dark spheres), medium repression (middle; dark spheres/blastoid-like structures) and no LTR5Hs repression (right; blastoid-like structures). Images are representative of 23 LTR5Hs-CARGO ( $n = 3$  biological replicates) and 24 nontarg-CARGO ( $n = 2$  biological replicates) clonal cell lines. Scale bar, 400  $\mu\text{m}$ . **f**, PCA of the bulk RNA-seq transcriptomes obtained from nontarg-CARGO and LTR5Hs-CARGO clonal cell lines with high or medium expression levels (total of 11 clonal cell lines in biological duplicates or triplicates). **g**, Representative images of blastoid (left;  $n = 11$  from three biological replicates) or dark sphere (right;  $n = 13$  from three biological replicates) immunostained with cleaved CASP3 (red) and DAPI (blue). Scale bars, 100  $\mu\text{m}$ . **h**, Quantification of cleaved CASP3 immunostainings described in panel **g**. Red diamonds display the median values. Significance was determined by an unpaired, two-tailed Student's *t*-test:  $P = 8.5 \times 10^{-7}$ .

datasets<sup>17,18</sup> revealed high HERVK expression in the epiblast and hypoblast lineages and low expression in the trophectoderm (Extended Data Fig. 1b–e).

We implemented the blastoid generation protocol<sup>17</sup> from hnPSCs, obtaining approximately 70% efficiency (Fig. 1a, Extended Data Fig. 1f and Supplementary Video 1). scRNA-seq analysis of the generated blastoids revealed transcriptome profiles consistent with the presence

of the three blastocyst lineages and absence of the post-implantation embryo lineages (Extended Data Fig. 1g,h). We also performed immunostainings with markers of the epiblast (KLF17, NANOG, SUSD2 and IFI16)<sup>18–20</sup>; the trophectoderm (GATA3)<sup>18,21</sup>; the hypoblast (SOX17 and GATA4)<sup>18,22</sup>; and the HERVK envelope (Fig. 1b, Extended Data Fig. 1i and Supplementary Table 1), all of which showed staining patterns consistent with human blastocysts<sup>9,19,21,22</sup>.



We previously reported that CARGO-CRISPRi allows for efficient and selective perturbation of HERVK LTR5Hs function across the genome<sup>10,23</sup>. In brief, a 12-mer guide RNA (gRNA) array was designed to target the majority of 697 LTR5Hs instances (referred to as LTR5Hs-CARGO), along with a control non-targeting array (nontarg-CARGO). To study the functional effect of HERVK LTR5Hs repression on human blastoid formation, we generated hnPSCs expressing a cumate-inducible catalytically dead version of Cas9 (dCas9) fused to the transcriptional repressor KRAB (KRAB-dCas9) and then introduced LTR5Hs-CARGO or nontarg-CARGO arrays to generate clonal cell lines (Fig. 1c). We confirmed that the induction of KRAB-dCas9 in LTR5Hs-CARGO hnPSCs but not in nontarg-CARGO hnPSCs resulted in H3K9me3 deposition across the majority of LTR5Hs instances and in the repression of LTR5Hs-originating transcripts (Extended Data Fig. 2a–c).

We next asked how LTR5Hs repression affects blastoid formation by inducing blastoid generation from 24 and 23 distinct nontarg-CARGO and LTR5Hs-CARGO clonal cell lines, respectively (Fig. 1c). We then measured blastoid formation efficiency as a function of LTR5Hs expression levels utilizing TaqMan probes (Fig. 1c–e and Supplementary Table 1). We observed a correlation between LTR5Hs expression and blastoid-forming potential, with high repression of LTR5Hs activity being incompatible with blastoid formation and instead resulting in structures resembling dark spheres. At intermediate levels of repression, blastoid-like structures still formed, albeit at reduced efficiencies, whereas lines with poor LTR5Hs repression formed blastoids at efficiencies comparable with the nontarg-CARGO hnPSCs. To rule out off-target effects, we cloned a new gRNA array, orthogonal to the LTR5Hs-CARGO array (named LTR5Hs-Ortho-CARGO) capable of repressing LTR5Hs and its selected gene targets (Extended Data Fig. 2d). In agreement with the LTR5Hs-CARGO, LTR5Hs-Ortho-CARGO hnPSCs also failed to form blastoids (Extended Data Fig. 2e).

HERVK retains coding capacity for viral proteins that can be detected in human embryos<sup>9</sup>. To explore whether loss of these viral proteins is responsible for the blastoid formation failure, we performed rescue experiments by genomic integration of an active transgene encoding the HERVK viral proteins gag, pro and pol<sup>24</sup>. This transgene failed to restore the blastoid formation capacity of high LTR5Hs repression hnPSCs, suggesting that HERVK viral proteins alone are not responsible for the dark spheres phenotype (Extended Data Fig. 2f). Together, these results indicate that LTR5Hs activity impacts blastoid formation and suggest a non-neutral contribution of HERVK to human pre-implantation.

## Gene misregulation upon LTR5Hs repression

The different blastoid formation capacity of high and medium LTR5Hs repression hnPSCs prompted us to analyse their underlying gene expression changes by RNA-seq after 96 h of repression (we note that hnPSCs cannot be maintained long term after induction of LTR5Hs repression). Principal component analysis (PCA) revealed that although all LTR5Hs-CARGO clones separated from the nontarg-CARGO clones, the medium repression clones clustered much closer to the controls than the high repression clones (Fig. 1f). Indeed, we confirmed a stronger misregulation of gene expression in high versus medium repression clones than nontarg-CARGO controls, in both the number of affected genes and the magnitude of the responses (Extended Data Fig. 2g,h). Gene Ontology of transcripts dysregulated in the high but not in the medium repression clones revealed categories of embryo morphogenesis and immune response among others, suggesting that the high-repressing clones undergo additional gene dysregulation (Supplementary Table 2).

The dark spheres obtained upon near-full LTR5Hs repression do not cavitate and look homogeneous under bright field (Fig. 1e). Bulk RNA-seq of the dark spheres compared with bulk RNA-seq of control blastoids detected a clear separation of these transcriptomes in the

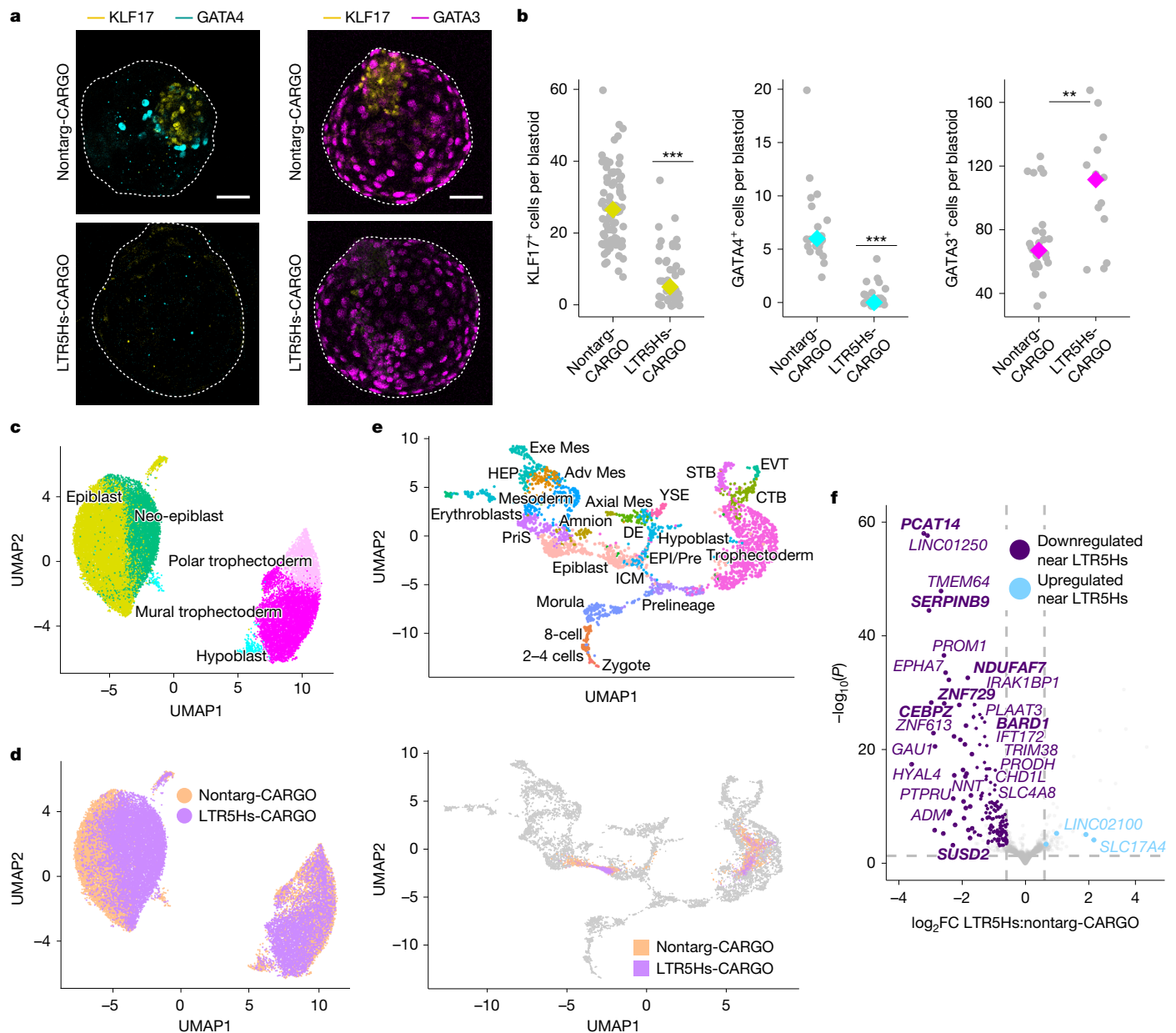
PCA space and identified differentially expressed genes (Extended Data Fig. 2i,j). Gene Ontology analysis of these transcripts revealed categories related to morphogenesis, migration, cell proliferation and others (Supplementary Table 2). Among the upregulated genes, some transcripts pointed to apoptosis (for example, *CASP7*), thus we systematically investigated whether apoptotic genes were differentially regulated. Indeed, comparison with a set of curated apoptosis genes<sup>25</sup> identified upregulation of apoptotic gene transcripts in the dark spheres (Extended Data Fig. 2k). To confirm this, we stained blastoids and dark spheres with the apoptotic marker cleaved CASP3. Although each blastoid displayed a median of three cleaved CASP3<sup>+</sup> cells, for dark spheres, the median was 29 (Fig. 1g,h). This demonstrates that under high LTR5Hs repression, hnPSCs undergo widespread gene expression changes incompatible with blastoid formation and consistent with an apoptotic phenotype.

## LTR5Hs repression affects lineage identity

We next asked whether hypomorphic blastoids with medium levels of LTR5Hs repression show lineage defects. Immunostainings with epiblast (KLF17 and *SUSD2*), hypoblast (*GATA4*) and trophectoderm (*GATA3*) markers revealed that medium LTR5Hs-CARGO blastoids had a diminished number of KLF17<sup>+</sup> and *GATA4*<sup>+</sup> cells, and a decreased *SUSD2* signal compared with nontarg-CARGO blastoids, suggesting defects in the epiblast and hypoblast lineages (Fig. 2a,b and Extended Data Fig. 3a). Of note, this decrease was not due to a loss of epiblast-like cells, as *NANOG* was detected in both nontarg-CARGO and LTR5Hs-CARGO blastoids (Extended Data Fig. 3b). By contrast, we detected an increased number of *GATA3*<sup>+</sup> cells, accompanied by smaller inner cell mass to trophectoderm ratios (Fig. 2a,b and Extended Data Fig. 3c).

To systematically interrogate changes in gene expression and lineage allocation associated with LTR5Hs repression, we profiled transcriptomes of single cells by particle-templated instant partition sequencing (PIP-seq; Extended Data Fig. 3d and Supplementary Table 3). We selected two medium repression lines (circles outlined in black, Fig. 1d) and two nontarg-CARGO lines. We generated a global embedding of all samples amounting to 31,028 cells and 37,468 genes and performed cell cluster annotation based on markers of the blastocyst<sup>18</sup> (Fig. 2c and Extended Data Fig. 3e). In parallel, we coloured cells based on their nontarg-CARGO or LTR5Hs-CARGO origin (Fig. 2d). Cells expressing markers of all blastocyst lineages were recovered in our analysis (Fig. 2c and Extended Data Fig. 3e). Within the trophectoderm lineage, we could distinguish mural and polar trophectoderm based on expression of *NR2F2*, *PGF* and *CYP19A1* in the latter<sup>18,26,27</sup> (Extended Data Fig. 3e). Projection of the transcriptomes into a collection of human embryo datasets (Methods) confirmed that the nontarg-CARGO cells matched transcriptomes of the pre-implantation embryo (Fig. 2e, orange). Increased resolution of clustering failed to call amnion or mesoderm clusters. In agreement, amnion and mesoderm markers were either not expressed or did not overlap a specific cluster (Extended Data Fig. 3f).

We observed the emergence of a new epiblast-adjacent cluster (dark green in Fig. 2c), which we refer to as ‘neo-epiblast’, by virtue of its transcriptional proximity to the epiblast. Although the neo-epiblast cells originated predominantly from the LTR5Hs-CARGO blastoids (purple in Fig. 2d and Extended Data Fig. 3g), the epiblast cluster was populated mostly by nontarg-CARGO cells, indicating that the neo-epiblast stems from gene expression changes driven by LTR5Hs repression. Projection of the LTR5Hs-CARGO cells into the human embryo reference datasets revealed they clustered with a less mature pre-implantation epiblast than the nontarg-CARGO epiblast-like cells (Fig. 2e and Extended Data Fig. 3h), suggesting that LTR5Hs repression affects epiblast maturation in blastoids. Nonetheless, HERVK LTR5Hs repression is not promoting a more pluripotent state, as the only naive pluripotency transcription factors changing upon LTR5Hs repression are *DPPA5* (higher expression) and *KLF17* (lower expression). We also detected a minor decrease



**Fig. 2 | LTR5Hs activity is required for proper lineage acquisition in blastoids.**

**a**, Representative confocal images of nontarg-CARGO and LTR5Hs-CARGO blastoids stained with lineage-specific antibodies (KLF17 (an epiblast marker) is in yellow,  $n = 78$  for nontarg-CARGO and  $n = 41$  for LTR5Hs-CARGO; GATA4 (a hypoblast marker) is in cyan,  $n = 22$  for nontarg-CARGO and  $n = 24$  for LTR5Hs-CARGO; GATA3 (a trophoblast marker) is in magenta,  $n = 27$  for nontarg-CARGO and  $n = 14$  for LTR5Hs-CARGO). Stained blastoids represent  $n \geq 4$  biological replicates. Scale bars, 50  $\mu\text{m}$ .

**b**, Counting of cells showing positive staining for the indicated markers in the nontarg-CARGO and LTR5Hs-CARGO blastoids as detailed in panel **a**. The grey dots represent the number of cells in individual blastoids. The yellow, cyan or magenta diamonds represent the median. Counted blastoids were generated in  $n \geq 4$  biological replicates; specific numbers counted are as follows: KLF17,  $n = 78$  for nontarg-CARGO and  $n = 41$  for LTR5Hs-CARGO; GATA4,  $n = 22$  for nontarg-CARGO and  $n = 24$  for LTR5Hs-CARGO; and GATA3,  $n = 27$  for nontarg-CARGO and  $n = 14$  for LTR5Hs-CARGO. Significance was determined using an unpaired, two-tailed Student's  $t$ -test for KLF17 ( $P = 2.2 \times 10^{-16}$ ), GATA4 ( $P = 3.69 \times 10^{-10}$ ) and GATA3 ( $P = 0.001$ ).

**c, d**, Uniform manifold and approximation projection (UMAP) of transcriptomes from single cells dissociated from nontarg-CARGO and LTR5Hs-CARGO

blastoids. The colours indicate cells belonging to the same lineage-specific clusters, as specified (**c**), or indicate the genotype of origin (orange for nontarg-CARGO blastoids and purple for LTR5Hs-CARGO blastoids; **d**).

**e**, UMAP of a reference collection of human embryo scRNA-seq transcriptomes<sup>51</sup> (top) and projection of nontarg-CARGO (orange) or LTR5Hs-CARGO (purple) PIP-seq results in that UMAP (bottom). Adv Med, advanced mesoderm; Axial Mes, axial mesoderm; CTB, cytotrophoblast; DE, definitive endoderm; EPI/Pre, epiblast prelineage; EVT, extravillous trophoblast; Exe Mes, extraembryonic mesoderm; HEP, haemato-endothelial progenitors; ICM, inner cell mass; PriS, primitive streak; STB, syncytiotrophoblast; YSE, yolk sac endoderm.

**f**, Volcano plot representing gene expression changes in LTR5Hs-CARGO versus nontarg-CARGO blastoid epiblast cells (using combined cells from epiblast and neo-epiblast clusters). Purple and light blue represent genes within 250 kb of an LTR5Hs, and the grey dots represent any other gene. Bold indicates genes further explored in Fig. 3b. The vertical dashed lines indicate statistically significant genes, fold change (FC)  $< -1.5$  or  $FC > 1.5$ . The horizontal dashed line indicates  $-\log_{10}(P) = 0.05$ . DESeq2 Wald test false discovery rate (FDR) of 5%.  $P$  values were corrected for multiple testing using the Benjamini-Hochberg method<sup>52</sup>.

in the primed pluripotency markers, consistent with the neo-epiblast cells being more immature (Extended Data Fig. 4a–c).

In addition to changes in the epiblast, we also observed a reduced allocation of LTR5Hs-CARGO cells to the hypoblast cluster, confirmed by the projections into the human embryo reference datasets (Fig. 2c–e and Extended Data Fig. 3g,h). The trophoctoderm was the least affected by LTR5Hs repression overall, in agreement with its lowest HERVK expression. However, in contrast to the immunostainings, we did not detect an increased proportion of trophoctoderm cells in the scRNA-seq analysis of the LTR5Hs-CARGO blastoids (see note in Methods). Beyond this caveat, we noticed a diminished contribution of the LTR5Hs-CARGO cells to the polar trophoctoderm cluster and a decreased expression of *IL6*, a highly expressed interleukin in the polar trophoctoderm that signals to the epiblast (Fig. 2c,d and Extended Data Fig. 4d). As maturation of the polar trophoctoderm is dependent on the signals from the epiblast<sup>17</sup>, its decrease in the LTR5Hs-CARGO blastoids may be an indirect consequence of the defective epiblast.

To test whether lineage defects were recapitulated in 2D hnPSCs cultures, we analysed the effect of LTR5Hs repression on direct trophoctoderm and hypoblast differentiations. LTR5Hs repression did not change the potential of hnPSCs to differentiate into the trophoctoderm (TROP2<sup>+</sup> cells; Extended Data Fig. 5a and quantified in Extended Data Fig. 5b) or the hypoblast (ANPEP<sup>+</sup> cells; Extended Data Fig. 5c and quantified in Extended Data Fig. 5d). Thus, the hypoblast defect that we observed in blastoids may be an indirect consequence of perturbed epiblast function or maturation. Overall, our results show that LTR5Hs repression disrupts lineage determination in human blastoids and results in a major change in the blastoids epiblast transcriptome.

### LTR5Hs regulates epiblast transcription in *cis*

To investigate transcriptome changes in the blastoids, we performed differential gene expression analysis between the nontarg-CARGO and the LTR5Hs-CARGO epiblast-like cells (Methods). We identified 255 and 87 transcripts that were downregulated and upregulated, respectively (Extended Data Fig. 5e and Supplementary Table 4). Only 122 and 24 of these genes significantly changed in the high and medium repression hnPSCs, respectively, highlighting distinct regulatory outputs of LTR5Hs in 2D hnPSCs cultures versus blastoids. LTR5Hs elements function as enhancers over distances of up to 250 kb from their target promoters<sup>10</sup>. We found that 84% of the downregulated transcripts in the epiblast lie within 250 kb of an LTR5Hs, compared with 1.62 Mb of median distance from all human protein-coding genes. This suggests direct regulation by LTR5Hs in *cis* (Fig. 2f). As expected, downregulated genes also display a decrease in expression when comparing neo-epiblast and epiblast clusters (Extended Data Fig. 5f).

The upregulated genes were typically not near LTR5Hs, suggesting indirect effects (Fig. 2f). However, many upregulated genes are associated with trophoctoderm and/or placental development (Extended Data Fig. 5g). As examples, *CSH2* encodes for the lactogen hormone<sup>28</sup>; *GATA2* is a trophoctoderm transcription factor<sup>18</sup>; *TACSTD2* and *DAB2* are trophoblast markers<sup>18,29</sup>; and, notably, *TGFA* increases blastocoel size in mouse<sup>30</sup>, suggesting that it could contribute to the larger trophoctoderm compartment in the LTR5Hs-CARGO blastoids (Fig. 2a,b and Extended Data Fig. 5g). Together, these observations point to LTR5Hs elements regulating the blastoid epiblast genes in *cis*. This is in turn associated with an indirect upregulation of trophoctoderm specification and maturation genes. Nonetheless, despite this upregulation, these cells remain closer to the naive pluripotent state in their overall transcriptome (Fig. 2e and Extended Data Fig. 4b).

### LTR5Hs elements are epiblast enhancers

We sought to address whether the effects on the epiblast gene expression were dependent on the LTR5Hs DNA sequence. To establish

sequence dependency and rule out ectopic silencing in our KRAB–dCas9 experiments, we selected six LTR5Hs elements located in *cis* to genes downregulated in the neo-epiblast with potential or known functions in development (Extended Data Fig. 5f). For example, *SUSD2* is a marker of human naive pluripotency<sup>20,31,32</sup>, *ZNF729* of naive and formative pluripotency<sup>32</sup>, and *BARD1* is a BRCA1 partner essential for embryonic development in mice<sup>33</sup>.

We engineered hnPSCs in which each of the six LTR5Hs elements has been homozygously or heterozygously deleted, one at a time, in multiple clonal lines (Fig. 3a). We then measured the effects of each deletion on gene expression compared with wild-type cells that underwent clonal selection in parallel (oligos in Supplementary Table 1). In each case, we observed that LTR5Hs deletion led to downregulation of the candidate target gene (Fig. 3b). For *SERPINB9*, homozygous deletions could not be recovered. We also noticed a slower growth of the  $\Delta$ LTR5Hs *ZNF729* clones, suggesting a role in proliferation. Apart from this, the clonal cell lines were morphologically undistinguishable from wild-type hnPSCs (Extended Data Fig. 6a).

In some cases, LTR5Hs deletion affected the expression of more than one gene (for example, *SUSD2/CABIN1*, *SERPINB9/SERPINB6* and *CEBPZ/NDUFAF7*; Fig. 3b), whereas in others, there was no effect on other genes located in *cis* (Extended Data Fig. 6b), suggesting selectivity in promoter responsiveness. For each deletion, we observed at least one unaffected gene, confirming that the deletions are not broadly altering the locus. Overall, promoters downregulated upon LTR5Hs deletions were located at various distances from LTR5Hs and represented different LTR5Hs–promoter orientations. Considering that dependence on DNA sequence, the ability to activate distally and independence of orientation are hallmarks of enhancers, we conclude that LTR5Hs elements function as hominoid-specific enhancers for many human epiblast genes.

### Hominoid transcriptome diversification

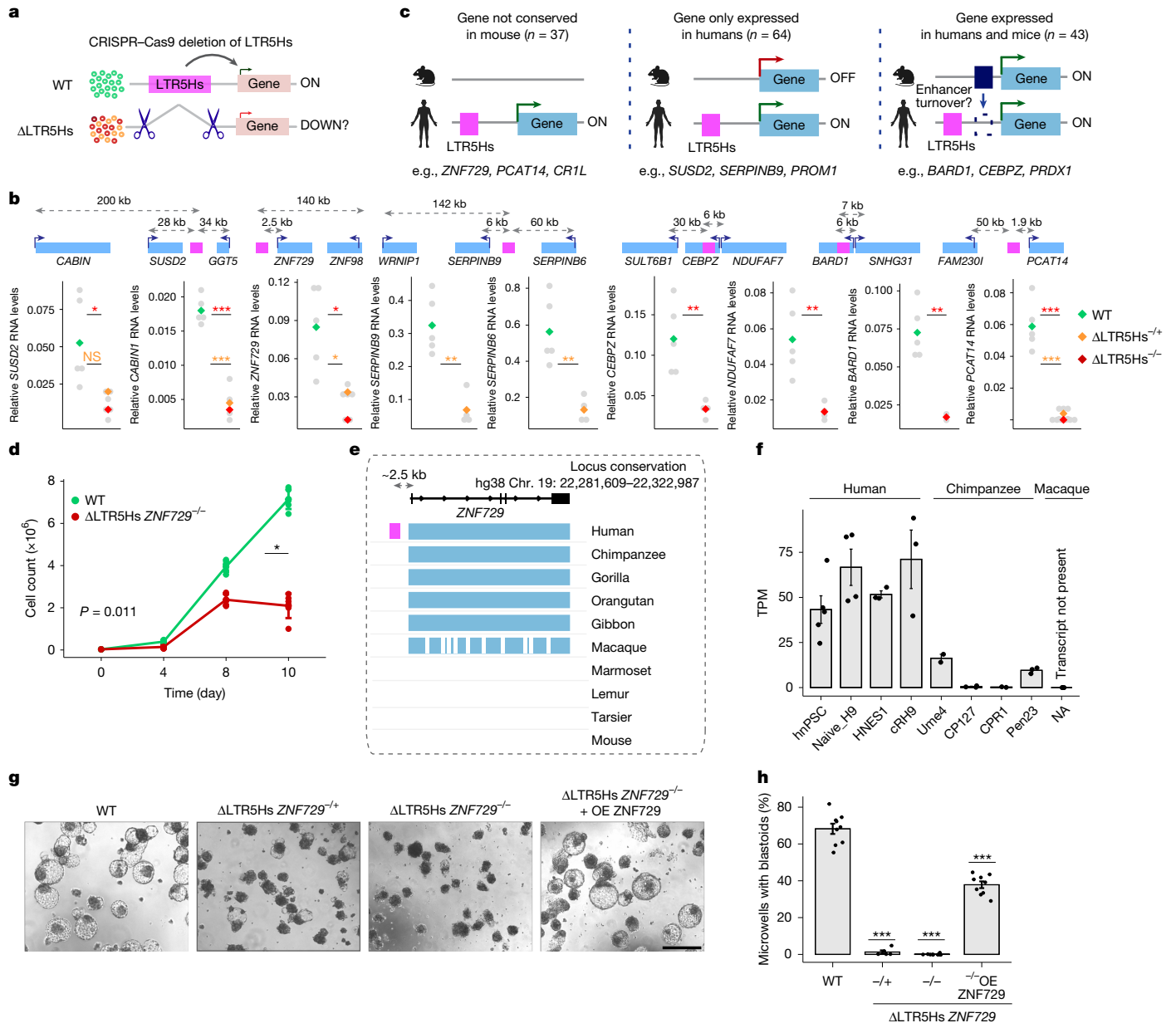
We next asked whether LTR5Hs had a role in the hominoid-specific diversification of the epiblast transcriptome. We investigated the expression and conservation of LTR5Hs-regulated genes between humans, marmosets (a primate that lacks HERVK LTR5Hs) and mice. We focused on candidate direct gene targets of LTR5Hs, defined as those downregulated upon LTR5Hs repression in the epiblast of blastoids and located within 250 kb of an LTR5Hs. Then, we drew on the transcriptomes of staged-match human, marmoset and mouse pre-implantation embryos<sup>34</sup>. Out of the 144 LTR5Hs direct target genes, 37 lack an orthologue in mice and thus are not expressed, whereas 29 do not have an orthologue in marmoset (Fig. 3c, Extended Data Fig. 7a and Supplementary Table 5).

The remaining genes have a mouse or marmoset orthologue, and we analysed their expression in the pre-implantation epiblast of these species, classifying the genes as ‘expressed’ or ‘not expressed’ (Methods). Among the conserved genes, 64 were expressed in humans but not in mice, whereas 53 were expressed in humans but not in marmoset (Fig. 3c and Extended Data Fig. 7a). Thus, the *cis*-regulatory activity of LTR5Hs has contributed to the species-specific diversification of the epiblast transcriptome, with a 101 and 82 LTR5Hs-dependent genes being expressed in humans but not in mice or marmosets, respectively.

### An essential human-specific LTR5Hs insertion

Other orthologous genes showed conserved epiblast expression, despite being dependent on LTR5Hs in humans. This may reflect ‘transposon addition’<sup>35,36</sup>, in which insertion of the strong LTR5Hs enhancer in hominoids may have led to turnover of the ancestral enhancer (or enhancers). Accordingly, we speculated that the LTR5Hs effect on development is more likely to arise from such transposon addition than from young LTR5Hs elements regulating recently emerged genes.





**Fig. 3 | A human-specific LTR5HS insertion is essential for blastoid formation.**

**a**, Strategy for testing LTR5HS sequence dependency. WT, wild-type. **b**, RT-qPCR results of the indicated genes in wild-type or  $\Delta$ LTR5HS hnPSCs. RNA values were normalized to *RPL13A*. Above each plot, schematics of the locus are included. The grey dots represent expression in each clone, and the diamonds represent median values. *n* (independent clones) per gene: *SUSD2/CABIN*, WT = 5, heterozygous = 2 and homozygous = 5; *ZNF729*, WT = 5, heterozygous = 4 and homozygous = 2; *SERPINB9/6*, WT = 5 and heterozygous = 4; *CEBPZ/NDUFAF7*, WT = 5 and homozygous = 4; *BARD1*, WT = 5 and homozygous = 2; and *PCAT14*, WT = 5, heterozygous = 6 and homozygous = 6. Significance was determined by an unpaired, two-tailed Student's *t*-test. The asterisk colour indicates homozygous (red) or heterozygous (yellow); specific *P* values are included as source data. NS, not significant. **c**, Expression and conservation of the LTR5HS-regulated genes between humans and mice using data in ref. 34. The silhouettes of the human and mouse were created in BioRender. Fueyo, R. (2025) <https://BioRender.com/tmgd0pn>. **d**, Cell proliferation curves in wild-type (green) and  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup> hnPSCs (red). *n* = 3 biological replicates. Data represent mean  $\pm$  s.d. Significance was determined by one-way analysis of variance

(ANOVA). **e**, Cartoon displaying the conservation of the gene *ZNF729* and its regulatory landscape across species. The LTR5HS element is only present in humans. Data adapted from Cactus UCSC tracks<sup>33,34</sup>. Chr., chromosome. **f**, *ZNF729* expression in human<sup>29,55,56</sup> and chimpanzee<sup>39</sup> naive PSCs; rhesus monkey naive PSCs lack the transcript. Data are presented as mean  $\pm$  s.e.m. Each dot represents a bulk RNA-seq biological replicate: hnPSC = 4, naive\_H9 = 4, HNES1 = 3, cRH9 = 3, Ume4 = 2, CP127 = 3, CPR1 = 2 and Pen23 = 3. TPM, transcripts per million. **g**, Representative bright-field images of blastoids or dark spheres generated from wild-type hnPSCs,  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup>,  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup> hnPSCs or  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup> hnPSCs expressing a *ZNF729* transgene (rescue,  $\Delta$ LTR5HS *ZNF729* + overexpression (OE) *ZNF729*). Scale bar, 400  $\mu$ m. **h**, Quantification of blastoid formation efficiency from panel *g*. Data are presented as mean  $\pm$  s.e.m. Each dot represents a biological replicate from one clone; we used three clones per condition in biological triplicates. Significance was determined by an unpaired, two-tailed Student's *t*-test: *P* =  $7.6 \times 10^{-10}$  for *ZNF729*<sup>-/-</sup>, *P* =  $4.47 \times 10^{-14}$  for  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup> and *P* =  $7.63 \times 10^{-9}$  for  $\Delta$ LTR5HS *ZNF729* + OE *ZNF729*.

Contrary to this expectation, the only LTR5HS deletion with an observable hnPSC phenotype was at the *ZNF729* locus. Indeed,  $\Delta$ LTR5HS *ZNF729*<sup>-/-</sup> hnPSCs exhibited longer doubling times (34 h versus 19 h

for wild-type cells; Fig. 3d). This was surprising, as this LTR5HS insertion is unique to humans, and the *ZNF729* gene is also evolutionary young (Fig. 3e). Although the precise evolutionary age of KRAB zinc-finger

genes is challenging to determine, previous analyses suggest that *ZNF729* emerged in the Old World anthropoids (*Catarrhini*) lineage<sup>37,38</sup> (Supplementary Table 5).

In humans, *ZNF729* expression is robust and correlates with HERVK activity (Fig. 3f and Extended Data Fig. 7b). However, although *ZNF729* is present in the chimpanzee genome, it is poorly expressed in naive PSCs in the chimpanzee<sup>39</sup>. Moreover, unguided transcriptome assembly (Methods) uncovered no evidence for *ZNF729* expression in naive PSCs in the macaque<sup>40</sup>. This suggests that *ZNF729* expression during pre-implantation may be unique to humans and linked to the insertion of the LTR5Hs enhancer. Of note, in agreement with enhancer rather than alternative promoter function, this LTR5Hs element displays enhancer chromatin marks (high H3K27ac and low H3K4me3 levels) and *ZNF729* transcript reads originate exclusively from the promoter according to the long-read human embryo transcriptomes (Extended Data Fig. 7c,d).

Clonal cell lines lacking the LTR5Hs *ZNF729* element in heterozygosity ( $\Delta$ LTR5Hs *ZNF729*<sup>+/−</sup>) or homozygosity ( $\Delta$ LTR5Hs *ZNF729*<sup>−/−</sup>) failed to form blastoids and instead gave rise to dark spheres resembling those observed on en masse LTRHs repression (Fig. 3g,h and Extended Data Fig. 7e). These dark spheres also expressed cleaved CASP3, as well as NANOG and GATA3 at the edges (Extended Data Fig. 7f–h). Blastoid potential of the  $\Delta$ LTR5Hs *ZNF729*<sup>−/−</sup> hnPSCs was partially rescued by a transgene encoding *ZNF729* cDNA, indicating that *ZNF729* mediates the phenotype (Fig. 3g,h, Extended Data Fig. 7e,i and Supplementary Fig. 1). The incomplete rescue is probably due to suboptimal transgene expression, rather than the regulation of multiple genes by the LTR5Hs insertion, as transcription of other genes at the locus is not affected by the LTR5Hs deletion (Extended Data Fig. 7j). Our results indicate that even highly species-specific retrotransposons can contribute to developmentally essential functions.

### **ZNF729 binds to GC-rich sequences in hnPSCs**

Structurally, *ZNF729* consists of a KRAB repressor domain and the highest number of zinc-finger domains in the human proteome: 37 (Fig. 4a). In addition, *ZNF729* is the most downregulated KRAB zinc-finger protein (KZFP) in the epiblast of the LTR5Hs-CARGO blastoids (Extended Data Fig. 8a). To acutely perturb *ZNF729* function, we derived homozygous clonal cell lines with *ZNF729* endogenously tagged with a dTAG-inducible degron and two haemagglutinin (HA) tags (Fig. 4a). We confirmed that upon dTAG<sup>−</sup>1 addition, *ZNF729*–FKBP–HA (*ZNF729*–FH) was degraded (Fig. 4b and Supplementary Fig. 1). HA chromatin immunoprecipitation followed by sequencing (ChIP-seq) of *ZNF729*–FH in control and dTAG<sup>−</sup>1-treated conditions identified 46,398 bound regions; more than 95% of these peaks were lost with dTAG<sup>−</sup>1 treatment, indicating specificity (Fig. 4c,h and Supplementary Table 6).

KZFPs bind to and repress transposable elements via recruitment of TRIM28 (also known as KAP1) and epigenetic repressors<sup>41</sup>. However, unlike most KZFPs, *ZNF729* binds mostly to non-repetitive DNA, with almost 46% of the peaks overlapping promoters (Fig. 4d). Within transposable elements, *ZNF729* binds to young transposable elements such as SVA\_D/F/B or LTR5Hs itself (Extended Data Fig. 8b,c). Motif discovery analysis at the non-repetitive peaks varied among tools, but all recovered motifs rich in G and C, with diverse spacing and different lengths (SeqPos examples in Fig. 4e and Supplementary Table 7). We wondered whether the binding of *ZNF729* at promoters can be explained by their high GC content in mammals. Indeed, GC content was much higher at *ZNF729*-bound than at unbound promoters (Fig. 4f). The discovery of GC-rich motifs could be a consequence, rather than a cause of *ZNF729* binding at the promoters. However, we also observed binding of *ZNF729* at simple repeats rich in G and/or C, even in the absence of association with promoters (Fig. 4g), suggesting a propensity of *ZNF729* for binding GC-rich sequences.

We next tested whether the KZFP partner TRIM28 colocalizes with *ZNF729*. ChIP-seq of TRIM28 in control *ZNF729*–FH hnPSCs confirmed colocalization of both proteins genome wide (Fig. 4h,i). Noticeably, repetitive regions exhibited higher TRIM28 levels compared to promoters, consistent with multiple KZFPs probably binding and recruiting TRIM28 to repeats (Fig. 4i and Extended Data Fig. 8c). Degradation of *ZNF729* affected TRIM28 genomic binding, diminishing the number of peaks from 17,460 to 3,160. This impact was less severe at repetitive regions, again consistent with potential KZFPs redundancy. By contrast, TRIM28 binding was strongly diminished at promoters, and TRIM28 regions not overlapping with *ZNF729* were unaffected (Fig. 4i). This indicates that *ZNF729* recognizes GC-rich sequences, including thousands of promoters where it recruits TRIM28.

### **ZNF729 controls basic cellular functions**

To investigate whether *ZNF729* regulates gene expression, we treated *ZNF729*–FH hnPSCs with dTAG<sup>−</sup>1 for 3 h or 24 h and performed RNA-seq. Differential gene expression analysis at 3 h detected mostly gene downregulation (Fig. 5a). Even at 24 h, there was a larger proportion of downregulated genes (Fig. 5b). Transposable element expression was not affected by loss of *ZNF729* (Extended Data Fig. 9a), consistent with potential KZFP redundancy and persistence of TRIM28 at transposable elements (Fig. 4i).

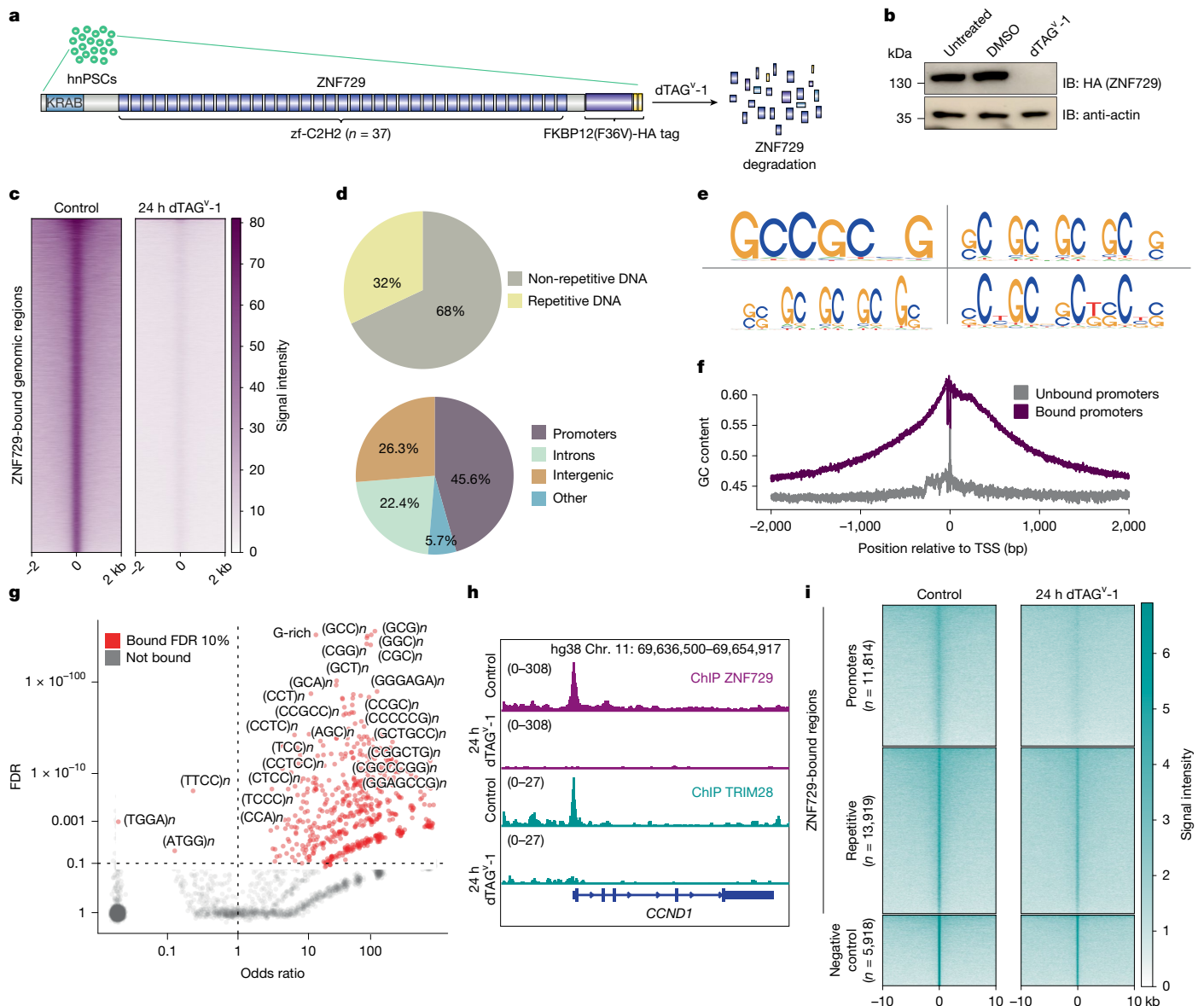
We next asked whether transcriptomic changes occurred at *ZNF729*-bound promoters. We defined directly regulated genes as those significantly misregulated upon loss of *ZNF729* and bound by *ZNF729* within −1 kb or +200 bp from the transcription start site. Of downregulated and upregulated promoters, 36% and 35%, respectively, were directly bound, indicating that *ZNF729* can act as a transcriptional activator or repressor (Extended Data Fig. 9b,c). TRIM28 occupancy at these promoters upon depletion of *ZNF729* shows a clear reduction of the TRIM28 signal, albeit the shapes of the profiles differ between the *ZNF729*-activated and *ZNF729*-repressed targets, suggesting that different protein complexes may be driving these distinct outputs (Extended Data Fig. 9d).

Housekeeping promoters are typically GC rich and highly and broadly active. *ZNF729*-bound promoters are GC rich and also display high expression (Extended Data Fig. 9b). Ontology analysis of genes regulated by *ZNF729* at 3 h are associated with basic cellular processes such as cell division, GTPase activity, RNA metabolic processes and others (Supplementary Table 8). This led us to reason that the  $\Delta$ LTR5Hs *ZNF729*<sup>−/−</sup> slow growth could be explained by cell-cycle genes affected upon depletion of *ZNF729*. Indeed, classic cell-cycle regulators such as *MYC*, *CDK1* or *CCND1* are bound and activated by *ZNF729* (Figs. 4h and 5c). These results suggest that *ZNF729* may be a major driver of phenotypes observed upon global LTR5Hs repression. Consistently, PCA revealed similarities between transcriptomes of *ZNF729*-depleted hnPSCs and those of high LTR5Hs repression dark spheres (Extended Data Fig. 9e).

### **Discussion**

Our results demonstrate how an evolutionary young gene, encoding a KZFP, gained expression in the human lineage upon capturing a retrovirus-derived enhancer. The product of the gene subsequently influenced essential cellular programs by binding and regulating GC-rich promoters (Fig. 5d).

We have shown that HERVK LTR5Hs activity is required for blastoid formation and lineage identity. LTR5Hs elements function as enhancers, and consequently, LTR5Hs repression severely affects the transcriptome of the epiblast of blastoids. LTR5Hs repression also leads to diminished hypoblast and polar trophectoderm cell numbers. Specification of these lineages in the embryo depends on the epiblast<sup>17,22</sup>, suggesting that these effects may originate in the defective epiblast. Alternatively,



**Fig. 4 | Widespread binding of ZNF729 at GC-rich sequences and promoters in hnPSC.** **a**, Schematic depicting endogenous tagging of ZNF729, with an FKBPV(F36V) and two HA tags in hnPSCs (ZNF729-FH hnPSCs). Upon addition of dTAG<sup>v</sup>-1, ZNF729 is degraded. The protein structure was drawn with IBS 2.0 (ref. 57). **b**, Representative western blot of ZNF729-FH hnPSCs untreated, treated with DMSO as control or treated with dTAG<sup>v</sup>-1 for 24 h. The membrane was blotted with an antibody to the HA tag and with  $\alpha$ -actin as the loading control blotted on the same membrane. For gel source data, see Supplementary Fig. 1. **c**, Heatmap displaying ZNF729-FH ChIP-seq signal over its bound genomic regions (n = 46,398) in DMSO treated (control) and 24-h dTAG<sup>v</sup>-1-treated ZNF729-FH hnPSCs (n = 2 biological replicates). **d**, Pie charts displaying the genomic features of ZNF729-FH-occupied regions. The pie charts show peak distribution between repetitive or non-repetitive DNA (top) and between promoters, intergenic or intronic DNA regions (bottom). **e**, Top four DNA sequence motifs obtained when performing

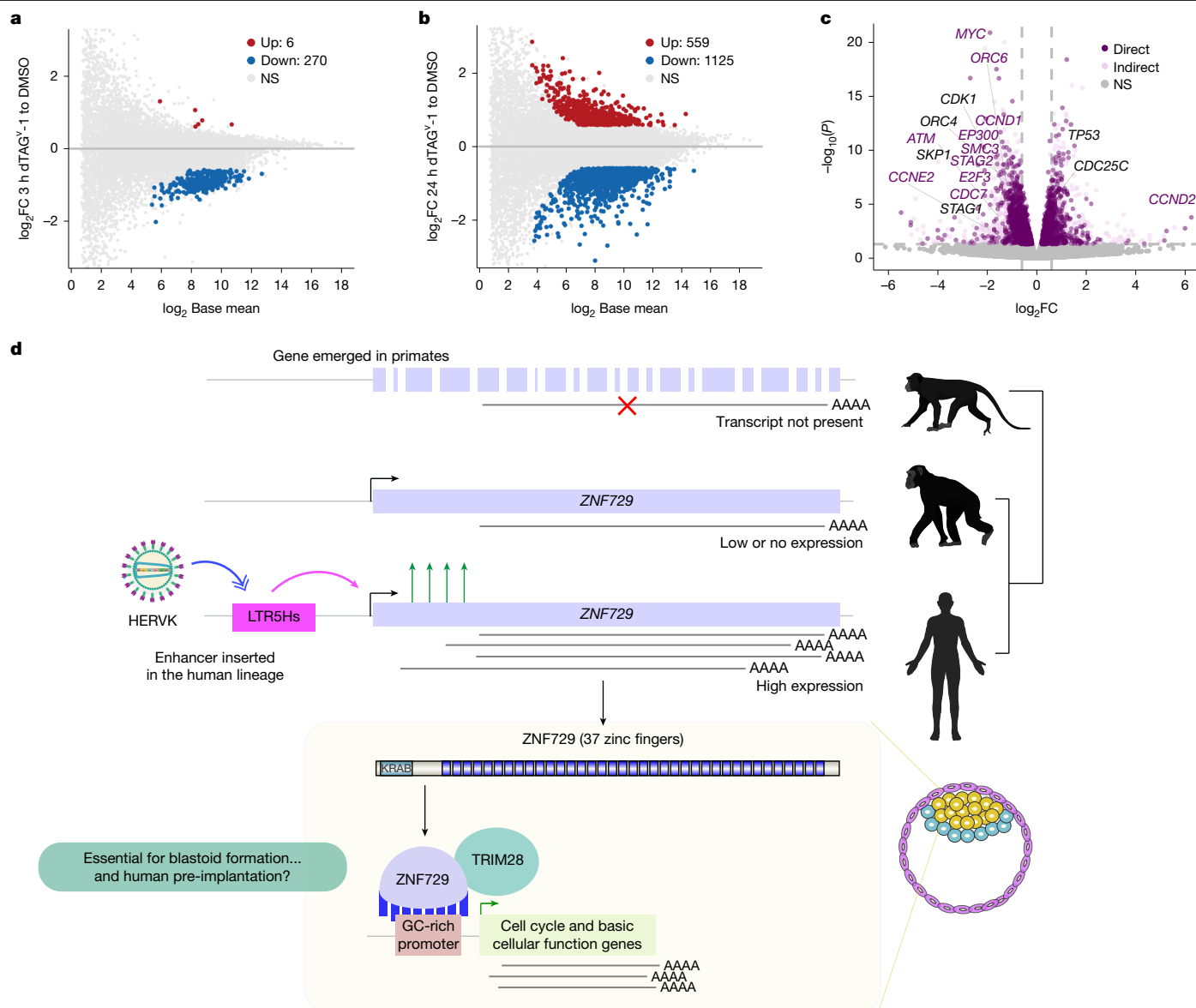
motif discovery analysis on the ZNF729-FH ChIP-seq non-repetitive peaks using SeqPos<sup>38</sup>. **f**, Plot representing GC content at ZNF729-FH-bound (purple line) or unbound (grey line) promoters. TSS, transcription start site. **g**, Scatter plot displays the odds ratio of ZNF729-FH binding to simple repeats that contain either G or C or both. The vertical dashed line indicates an odds ratio of 1. The horizontal dashed line indicates FDR of 10%. Significantly bound (FDR of 10%) are depicted in red and not bound in grey. **h**, IGV genome browser capture of the ZNF729-FH ChIP-seq signal (top two tracks, purple) or the TRIM28 ChIP-seq signal (bottom two tracks, turquoise) in DMSO-treated (control) and 24-h dTAG<sup>v</sup>-1-treated ZNF729-FH. **i**, Heatmap displaying the TRIM28 ChIP-seq signal over ZNF729-bound non-repetitive promoters (top), the repetitive regions (middle) or the regions bound by TRIM28 that do not overlap with ZNF729 (bottom). The ChIP-seq signals from control-treated (left) or dTAG<sup>v</sup>-1-treated (right) ZNF729-FH hnPSCs are shown (n = 2 biological replicates).

as described in mice, the hypoblast could be dedifferentiating back to the inner cell mass<sup>42</sup>.

The essentiality of HERVK LTR5Hs for blastoid formation probably stems from multiple mechanisms. One explored here in detail is the *cis*-regulatory activity of the LTR5Hs at the *ZNF729* locus, but other LTR5Hs enhancers and their target genes may also be required for human development. Moreover, a subset of HERVK LTR5Hs insertions

encodes retroviral proteins that can assemble viral-like particles in human blastocysts<sup>9,43</sup>. Although HERVK retroviral proteins do not rescue the failed blastoid formation, they may still have other vital roles in the embryo<sup>44</sup>. Of note, LTR10A and MER41B elements also exert *cis*-regulatory functions of placental genes in trophoblast stem cells, suggesting that other retrotransposon families may also be essential in embryogenesis<sup>45</sup>.





**Fig. 5 | ZNF729 is a transcriptional regulator of basic cellular functions in hnPSCs.** **a**, Logged intensity ratio (M) versus mean logged intensities (A) (MA) plot representing bulk RNA-seq results of ZNF729-FH hnPSCs treated with DMSO (control) or with dTAG<sup>v-1</sup> for 3 h ( $n = 4$ ). The y-axis shows gene expression fold changes in 3-h dTAG<sup>v-1</sup>-treated versus DMSO-treated ZNF729-FH hnPSCs. FDR of 5%. The horizontal line indicates fold change of 0. **b**, MA plot representing bulk RNA-seq results of ZNF729-FH hnPSCs treated with DMSO (control) or with dTAG<sup>v-1</sup> for 24 h ( $n = 4$ ). The y-axis shows gene expression fold changes in 24-h dTAG<sup>v-1</sup>-treated versus DMSO-treated ZNF729-FH hnPSCs. The horizontal line indicates fold change of 0. **c**, Volcano plot of gene expression changes in bulk

RNA-seq analyses from ZNF729-FH hnPSCs treated with DMSO (control) or with dTAG<sup>v-1</sup> for 24 h ( $n = 4$ ), coloured by the presence (dark purple) or absence (light pink) of ZNF729-FH binding at the gene promoter. The labelled dots represent genes involved in the cell cycle according to gene set enrichment analysis-curated gene sets. Significance was determined using DESeq2 Wald test, FDR of 5%.  $P$  values were corrected for multiple testing using the Benjamini-Hochberg method<sup>52</sup>. **d**, Model for human-specific gain of function of ZNF729-FH in early development, facilitated by the insertion of HERVK LTR5Hs. The silhouettes of the human, chimpanzee and macaque were created in BioRender. Fueyo, R. (2025) <https://BioRender.com/tmgd0pn>.

We revealed a requirement for the human-specific LTR5Hs *ZNF729* for proliferation and blastoid formation potential. *ZNF729* confers this essentiality by binding and regulating hundreds of crucial genes due to its affinity for GC-rich sequences, which are extremely abundant in mammalian housekeeping gene promoters. *ZNF729* contains an intact KRAB domain, capable of mediating repression in a screen of transcriptional effectors<sup>46</sup>. In agreement, we showed that *ZNF729* is the major transcription factor responsible for recruiting TRIM28 to promoters in hnPSCs. However, many regulated promoters are activated, rather than repressed, by *ZNF729*, an uncommon finding for a KZFP, albeit not a unique case<sup>47</sup>. Although the mechanism underlying this activation remains to be established, we note that TRIM28 has also been implicated

in gene activation through mediating RNA polymerase II (RNAPII) pause-release<sup>48–50</sup>. Alternatively, *ZNF729* activator function may be mediated by a heretofore unknown TRIM28-independent mechanism.

We wondered why an evolutionary young gene, controlled by a human-specific enhancer, is important for regulation of ancient cellular programs. We speculate that during primate evolution, the expansion of the zinc-finger array and mutations at amino acids contacting the DNA, ultimately resulted in a high affinity of *ZNF729* for GC-rich sequences. In turn, this targeted *ZNF729* to housekeeping promoters, potentially boosting their activity and providing a competitive proliferative advantage to cells expressing *ZNF729*. However, such regulation could have also led to a relaxation of the constraint on the ancient mechanisms

controlling expression of proliferation genes, with ZNF729 ultimately supplanting one of such mechanisms and becoming essential for preimplantation. These observations suggest that evolutionary remodeling of gene-regulatory networks can result not only in species-specific innovation but also create new dependencies and bestow essentiality on recently emerged *cis*-regulatory elements and genes.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-09571-1>.

- Shahbazi, M. N. & Pasque, V. Early human development and stem cell-based human embryo models. *Cell Stem Cell* **31**, 1398–1418 (2024).
- Hoyt, S. J. et al. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science* **376**, eabk3112 (2022).
- de Parseval, N. & Heidmann, T. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenet. Genome Res.* **110**, 318–332 (2005).
- Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-022-00457-y> (2022).
- Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
- Peaston, A. E. et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
- Sakashita, A. et al. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat. Genet.* **55**, 484–495 (2023).
- Modzelewski, A. J. et al. A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell* **184**, 5541–5558.e22 (2021).
- Grow, E. J. et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).
- Fuentes, D. R., Swigut, T. & Wysocka, J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7**, e35989 (2018).
- Pontis, J. et al. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**, 724–735.e5 (2019).
- Subramanian, R. P., Wildschutte, J. H., Russo, C. & Coffin, J. M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* **8**, 90 (2011).
- Shin, W. et al. Human-specific HERV-K insertion causes genomic variations in the human genome. *PLoS ONE* **8**, e60605 (2013).
- Zhu, M. & Zernicka-Goetz, M. Principles of self-organization of the mammalian embryo. *Cell* **183**, 1467–1478 (2020).
- Gerri, C., Menchero, S., Mahadevaiah, S. K., Turner, J. M. A. & Niakan, K. K. Human embryogenesis: a comparative perspective. *Annu. Rev. Cell Dev. Biol.* **36**, 411–440 (2020).
- Rugg-Gunn, P. J., Moris, N. & Tam, P. P. L. Technical challenges of studying early human development. *Development* **150**, dev201797 (2023).
- Kagawa, H. et al. Human blastoids model blastocyst development and implantation. *Nature* **601**, 600–605 (2022).
- Petropoulos, S. et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
- Lea, R. A. et al. KLF17 promotes human naive pluripotency but is not required for its establishment. *Development* **148**, dev199378 (2021).
- Brendenkamp, N., Stirparo, G. G., Nichols, J., Smith, A. & Guo, G. The cell-surface marker sushi containing domain 2 facilitates establishment of human naive pluripotent stem cells. *Stem Cell Rep.* **12**, 1212–1222 (2019).
- Gerri, C. et al. Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature* **587**, 443–447 (2020).
- Rood, M. et al. Human hypoblast formation is not dependent on FGF signalling. *Dev. Biol.* **361**, 358–363 (2012).
- Gu, B. et al. Transcription-coupled changes in nuclear mobility of mammalian *cis*-regulatory elements. *Science* **359**, 1050–1055 (2018).
- Lee, Y. N. & Bieniasz, P. D. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* **3**, e10 (2007).
- Mootha, V. K. et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
- Liu, D. et al. Primary specification of blastocyst trophectoderm by scRNA-seq: new insights into embryo implantation. *Sci. Adv.* **8**, eabj3725 (2022).
- Meistermann, D. et al. Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell* **28**, 1625–1640.e6 (2021).
- Josimovich, J. B., Atwood, B. L. & Goss, D. A. Luteotrophic, immunologic and electrophoretic properties of human placental lactogen. *Endocrinology* **73**, 410–420 (1963).
- Io, S. et al. Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023–1039.e13 (2021).
- Dardik, A. & Schultz, R. M. Blastocoel expansion in the preimplantation mouse embryo: stimulatory effect of TGF- $\alpha$  and EGF. *Development* **113**, 919–930 (1991).
- Messmer, T. et al. Transcriptional heterogeneity in naive and primed human pluripotent stem cells at single-cell resolution. *Cell Rep.* **26**, 815–824.e4 (2019).
- Kinoshita, M. et al. Capture of mouse and human stem cells with features of formative pluripotency. *Cell Stem Cell* **28**, 2180 (2021).
- McCarthy, E. E., Celebi, J. T., Baer, R. & Ludwig, T. Loss of Bard1, the heterodimeric partner of the Brca1 tumor suppressor, results in early embryonic lethality and chromosomal instability. *Mol. Cell. Biol.* **23**, 5056–5063 (2003).
- Boroviak, T. et al. Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, dev167833 (2018).
- Cosby, R. L., Chang, N.-C. & Feschotte, C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* **33**, 1098–1116 (2019).
- Chang, N.-C. et al. Gag proteins encoded by endogenous retroviruses are required for zebrafish development. *Proc. Natl Acad. Sci. USA* **122**, e2411446122 (2025).
- Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 (2017).
- Shao, Y. et al. GenTree, an integrated resource for analyzing the evolution and function of primate-specific coding genes. *Genome Res.* **29**, 682–696 (2019).
- Huang, T. et al. Inhibition of PRC2 enables self-renewal of blastoid-competent naive pluripotent stem cells from chimpanzee. *Cell Stem Cell* **32**, 627–639.e8 (2025).
- Fang, R. et al. Generation of naive induced pluripotent stem cells from rhesus monkey fibroblasts. *Cell Stem Cell* **15**, 488–497 (2014).
- Ecco, G., Imbeault, M. & Trono, D. KRAB zinc finger proteins. *Development* **144**, 2719–2729 (2017).
- Linneberg-Agerholm, M. et al. The primitive endoderm supports lineage plasticity to enable regulative development. *Cell* **187**, 4010–4029.e16 (2024).
- Bieda, K., Hoffmann, A. & Boller, K. Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. *J. Gen. Virol.* **82**, 591–596 (2001).
- Glaser, J. et al. Enhancer adoption by an LTR retrotransposon generates viral-like particles, causing developmental limb phenotypes. *Nat. Genet.* <https://doi.org/10.1038/s41588-025-02248-5> (2025).
- Frost, J. M. et al. Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat. Struct. Mol. Biol.* **30**, 527–538 (2023).
- Tycko, et al. High-throughput discovery and characterization of human transcriptional effectors. *Cell* **183**, 2020–2035.e16 (2020).
- Chen, W. et al. ZFP30 promotes adipogenesis through the KAP1-mediated activation of a retrotransposon-derived Pparg2 enhancer. *Nat. Commun.* **10**, 1809 (2019).
- Bacon, C. W. et al. KAP1 is a chromatin reader that couples steps of RNA polymerase II transcription to sustain oncogenic programs. *Mol. Cell* **78**, 1133–1151.e14 (2020).
- Yang, Y. et al. HIF-1 interacts with TRIM28 and DNA-PK to release paused RNA polymerase II and activate target gene transcription in response to hypoxia. *Nat. Commun.* **13**, 316 (2022).
- Bunch, H. et al. TRIM28 regulates RNA polymerase II promoter proximal pausing and pause release. *Nat. Struct. Mol. Biol.* **21**, 876–883 (2014).
- Zhao, C. et al. A comprehensive human embryo reference tool using single-cell RNA-sequencing data. *Nat. Methods* **22**, 193–206 (2025).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
- Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
- Kuderna, L. F. K. et al. Identification of constrained sequence elements across 239 primate genomes. *Nature* **625**, 735–742 (2024).
- Okubo, T. et al. Hypoblast from human pluripotent stem cells regulates epiblast development. *Nature* **626**, 357–366 (2024).
- Rostovskaya, M., Stirparo, G. G. & Smith, A. Capacitation of human naive pluripotent stem cells for multi-lineage differentiation. *Development* **146**, dev172916 (2019).
- Xie, Y. et al. IBS 2.0: an upgraded illustrator for the visualization of biological sequences. *Nucleic Acids Res.* **50**, W420–W426 (2022).
- Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

## Methods

### Ethics

This work was performed following the 2021 ISSCR Guidelines<sup>59</sup>. The use of blood-derived induced naive pluripotent stem cells for the experiments described in this article was approved by the Stanford Stem Cell Research Oversight committee (SCRO protocol number 900).

### Cell culture

Induced hnPSCs were generated from peripheral blood cells by over-expressing NANOG and KLF2 by Masaki et al.<sup>60</sup>; the official name of this cell line is PB004. Cells have been authenticated by STR. Cells were grown on irradiated Cf1 mouse embryonic fibroblast (MEF) feeder layers (A34181, Fisher Scientific). Before experiments entailing next-generation sequencing, hnPSCs were plated without MEFs (feeder-free conditions) using Geltrex (A1413301, Gibco) as a matrix. hnPSCs were grown in PXGL<sup>20</sup>. This medium consists of N2B27 as base, which is made by mixing 1:1 DMEM/F-12 (D8437, Sigma) and Neurobasal (21103049, Thermo) with the following added supplements: 2 mM L-glutamine (25030024, Thermo), 100  $\mu$ M 2-mercaptoethanol (M3148, Sigma), N2 and B27 supplements (17502048 and 7504044, Gibco) and 1 $\times$  antibiotic-antimycotic solution (A5955, Sigma-Aldrich). To make PXGL, we freshly supplemented the following chemicals: 1  $\mu$ M PD0325901 (S1036, Selleckchem), 2  $\mu$ M XAV939 (SM38-200, Cell Guidance Systems), 2  $\mu$ M Go 6983 (2285, Bio-Techne), 10 ng ml<sup>-1</sup> recombinant human LIF (300-05, Preprotech) and 1  $\mu$ g ml<sup>-1</sup> of doxycycline to sustain NANOG and KLF2 transgenes expression in hnPSCs. Doxycycline was eliminated from the media for the nontarg-CARGO or LTR5Hs-CARGO induction experiments. Cells were passaged using TrypLE Express (12-605-010, Fisher Scientific) every 3–4 days or whenever colonies were too confluent. The cell incubator was kept at 37 °C and humidified at 7% CO<sub>2</sub> and 5% O<sub>2</sub> (hypoxia). All cell lines were tested monthly for *Mycoplasma*. For KRAB-dCas9 and ZNF729 (rescue) inductions, we used 2 $\times$  water soluble cumate (QM150A-1, System Biosciences).

### Derivation of nontarg-CARGO and LTR5Hs-CARGO hnPSCs

hnPSCs were nucleofected using a Lonza 4D-Nucleofector using the P3 Primary X Kit-S (V4XP-3032, Lonza) with the DN100 program. Per nucleofection, we used 400,000 cells without MEF depletion. To generate the KRAB-dCas9 hnPSCs, 0.8  $\mu$ g of a piggyBac construct containing KRAB-dCas9 under a cumate-inducible promoter<sup>61</sup>, the cumate repressor CymR and a puromycin selection cassette were co-nucleofected with 0.2  $\mu$ g of the super piggyBac transposase (PB210PA-1, System Biosciences). Clones containing the integration were selected with puromycin (0.5  $\mu$ g ml<sup>-1</sup>) for three passages. KRAB-dCas9 hnPSCs cells were later nucleofected with 0.8  $\mu$ g of the piggyBac constructs containing nontarg-CARGO (#191319, Addgene)<sup>10</sup> or LTR5Hs-CARGO (#191316, Addgene)<sup>10</sup> and a neomycin selection cassette and 0.2  $\mu$ g of the super piggyBac transposase. Cells were then selected with 200  $\mu$ g ml<sup>-1</sup> of G418 for 10 days. Cells ( $n = 2,000$ ) were subsequently plated in a 10-cm<sup>2</sup> plate containing MEFs and fed every day. On days 8–9, sparse colonies were visible and were picked and expanded for the experiments. Cells were treated with puromycin and G418 every few passages to sustain proper KRAB-dCas9 and CARGO array expression, as we noticed these transgenes get silenced over the passages in hnPSCs. For the ‘orthogonal repression of LTR5Hs’ experiments, a distinct array of gRNAs targeting LTR5Hs was designed and cloned into piggyBac using CARGO<sup>23</sup> (gRNA sequences are in Supplementary Table 1). The LTR5Hs-Ortho-CARGO hnPSCs were generated as described above, with the only difference that this time the KRAB-dCas9 transgene was under a cumate-inducible *Efla* promoter to ensure high repression at the population level. Analysis of the role of the HERVK proteins in the dark spheres phenotype (‘rescue with HERVK ORFs’ experiment) was performed by selecting three high-repression LTR5Hs-CARGO clones that were previously demonstrated to give rise to dark spheres, and

integrating into them multiple copies of a piggyBac transgene encoding a tagBFP and the proteins gag, pro and pol<sup>24</sup> under a constitutive *Efla* promoter to ensure robust expression. High-repression LTR5Hs-CARGO hnPSCs positive for tagBFP were isolated and utilized for blastoid formation under cumate treatment to induce LTR5Hs repression.

### Genetic deletion of LTR5Hs and ZNF729 overexpression

Selected LTR5Hs elements were deleted from the genome using pairs of gRNAs designed using Benchling (Biology Software, 2023; Supplementary Table 1). crRNAs were purchased from IDT with the XT modification for stability. Cells ( $n = 400,000$ ) were nucleofected with a ribonucleoprotein complex containing 1.65  $\mu$ g of HiFi Cas9 Nuclease V3 (1081059, IDT) and 0.85  $\mu$ l of a 1:1 ratio of 100  $\mu$ M annealed tracrRNA and crRNA. Cells were passaged once and then 2,000 cells were plated on a 10-cm<sup>2</sup> plate with MEFs for colony picking. Clones were genotyped using PCR and Sanger sequencing, and heterozygous and homozygous clones were kept for experiments. For the rescue experiments in Fig. 3g,h, 400,000  $\Delta$ LTR5Hs *ZNF729*<sup>-/-</sup> hnPSCs were nucleofected with a piggyBac plasmid subcloned from a pcDNA3 vector, containing ZNF729-HA cDNA (purchased from Genscript) and a puromycin selection cassette. Super piggyBac transposase was co-nucleofected. Cells were selected with 0.5  $\mu$ g ml<sup>-1</sup> puromycin for 10 days and ZNF729-HA expression was tested by western blot.

### Derivation of ZNF729-FKBP(F36V)-HA hnPSCs

To endogenously tag *ZNF729*, we performed homology-directed repair at the locus with a donor DNA providing the FKBP(F36V)<sup>62,63</sup> and HA tags. To this end, we drew upon a previously published method<sup>64</sup> based on the combination of Cas9 ribonucleoproteins and delivery of the donor template by AAV6 viral vectors. To generate the AAV viral particles, 2  $\times$  15 cm<sup>2</sup> dishes of 293FT cells (R70007, Invitrogen) at 60% confluency were transfected. The day of transfection, the 293FT cells ‘complete cell media’ (DMEM/high-glucose medium (SH30243.FS, Cytiva), 10% FBS (100-106, GeminiBio), 1X non-essential amino acids (1114-0050, Gibco), 1X GlutaMAX (4109-0036, Gibco) and 1X antibiotic-antimycotic (1524-0062, Gibco)) was refreshed 6 h before transfecting. Transfection was carried out using 120  $\mu$ g polyethylenimine (PEI) per 15-cm<sup>2</sup> plate, 22  $\mu$ g of pDGM6 (#110660, Addgene)<sup>65</sup> and 6  $\mu$ g of AAV template (cloned in the pAAV-GFP backbone; #32395, Addgene)<sup>66</sup>. After 24 h, the medium was changed to ‘slow growth media’ (same as complete media, but with 2% FBS instead of 10%), and upon further 48 h of culture, the AAV viral particles were purified using one reaction of the AAVpro kit (6675, Takara Bio) and stored at -80 °C. The crRNA utilized to target the *ZNF729* C-terminal region was purchased from IDT with the XT modification for stability. Wild-type hnPSCs ( $n = 400,000$ ) were nucleofected with the ribonucleoprotein complex containing 1.65  $\mu$ g of HiFi Cas9 Nuclease V3 (1081059, IDT) and 0.85  $\mu$ l of a 1:1 ratio of 100  $\mu$ M annealed tracrRNA and crRNA. Cells were seeded in a plate containing MEFs, PXGL, the ROCK inhibitor Y-27632 and the AAV viral particles containing the donor template. Media were changed after 24 h. Cells were passaged once and then 2,000 cells were plated on a 10-cm<sup>2</sup> plate with MEFs for colony picking. Correct editing was analysed by PCR, Sanger sequencing and western blot. ZNF729 depletion was obtained upon addition of 500 nM of dTAG<sup>v</sup>-1 (6914, Tocris) for the indicated times.

### Blastoid formation

To generate blastoids, the protocol described in Kagawa et al.<sup>17,67</sup> was followed with minor changes. hnPSCs were grown on MEFs and dissociated the day of the experiment into single cells. MEFs were depleted by culturing the dissociated cells in PXGL over a gelatin matrix (G1393, Sigma-Aldrich) for 1 h. We used 24-well Aggrewell 400 (34415, Stem-Cell Technologies) plates as vessels. Upon multiple tests, we determined that starting from 76 cells per intended blastoid was optimal, so 91,200 hnPSCs were plated per well of the microwell plate (76  $\times$  1,200 microwells). On the day of plating, cells were cultured in N2B27



# Article

base medium containing 10  $\mu\text{M}$  Y-27632 (72304, StemCell Technologies). After 20–24 h, medium was changed to PALLY medium (N2B27 base medium supplemented with PD0325901 (1  $\mu\text{M}$ ), A83-01 (1  $\mu\text{M}$ ; HY-10432, MedChemExpress), 1-oleoyl lysophosphatidic acid sodium salt (LPA; 500 nM; 3854, Tocris), hLIF (10 ng  $\text{ml}^{-1}$ ) and Y-27632 (10  $\mu\text{M}$ )). PALLY medium was refreshed the next day. Seventy-two hours after plating, medium was replaced with medium containing 500 nM of LPA. At 96 h, structures were collected and analysed as needed. In those experiments in which 2 $\times$  cumate was added to induce KRAB–dCas9 expression, the drug was added on the day of plating concomitantly with cell aggregation.

## hnPSCs differentiation towards the trophoctoderm lineage

Trophoctoderm monolayer differentiation was completed as described previously<sup>29,68</sup>. In brief, hnPSCs were washed with PBS and then incubated with TrypLE Express for 10 min at 37 °C. Dissociated cells were washed in DMEM/F-12 (11-330-057, Thermo Fisher Scientific) with 0.1% Bovine Albumin Fraction V (15260037, Thermo Fisher Scientific) and resuspended in nTE-1 media (N2B27 media supplemented with 2  $\mu\text{M}$  PD325901, 2  $\mu\text{M}$  A83-01 and 10 ng  $\text{ml}^{-1}$  BMP4 (314-BP-010, R&D Systems)). Cells were counted and seeded to plates coated with 0.15  $\mu\text{g cm}^{-2}$  laminin511-E8 (AMS.892021, Amsbio) at a density of  $2 \times 10^4$  cells per  $\text{cm}^2$ . Twenty-four hours after plating, media were changed to nTE-2 media (N2B27 media supplemented with 2  $\mu\text{M}$  PD325901, 2  $\mu\text{M}$  A83-01 and 1  $\mu\text{g ml}^{-1}$  JAK inhibitor I (74022, StemCell Technologies)). Forty-eight hours after plating, the media were again changed to fresh nTE-2 media. To repress LTR5Hs elements during the differentiation, media were supplemented with 2 $\times$  water-soluble cumate. Differentiations took place under hypoxic conditions.

## hnPSCs differentiation towards the hypoblast lineage

Hypoblast monolayer differentiation from hnPSCs was completed as previously described<sup>55</sup>. In brief, hnPSCs were washed with PBS and then incubated with TrypLE Express for 10 min at 37 °C. Dissociated cells were washed in DMEM/F-12 with 0.1% Bovine Albumin Fraction V and resuspended in a six-factor ‘6 F media’ (N2B27 media supplemented with 25 ng  $\text{ml}^{-1}$  FGF4 (100-31, PeproTech; stabilized with 1  $\mu\text{g ml}^{-1}$  heparin sodium), 10 ng  $\text{ml}^{-1}$  recombinant human BMP4, 10 ng  $\text{ml}^{-1}$  recombinant human PDGF-AA (100-13A, Peprotech), 1  $\mu\text{M}$  XAV939 (SM38-10, Cell Guidance Systems), 3  $\mu\text{M}$  A83-01 (HY-10432, MedChem Express) and 0.1  $\mu\text{M}$  retinoic acid (R2625, Sigma-Aldrich). Cells were counted and seeded to plates coated with 0.15  $\mu\text{g cm}^{-2}$  laminin511-E8 at a density of  $5 \times 10^4$  cells per  $\text{cm}^2$ . Twenty-four hours after plating, the medium was replaced with fresh 6F media. Forty-eight hours after plating, the medium was changed to a seven-factor ‘7F media’, which includes the same factors used in the 6F media, with the addition of 10 ng  $\text{ml}^{-1}$  recombinant human IL-6 (200-06, PeproTech). To repress LTR5Hs elements during the differentiation, media were supplemented with 2 $\times$  water-soluble cumate. Differentiations took place under hypoxic conditions and flow cytometry measures were taken on day 3.

## Flow cytometry

After 3 days of trophoctoderm or hypoblast differentiation, 200,000 cells were used for staining. Cells were pelleted and resuspended in 100  $\mu\text{l}$  of N2B27 supplemented with 10  $\mu\text{M}$  Y-27632 and either a 1:100 dilution of TACTSD2-BV421 for the trophoctoderm differentiations (563243, BD Biosciences) or 1:200 dilution of ANPEP-BV421 for the hypoblast differentiations (301716, BioLegend). Cells were incubated on ice in the dark for 1 h and then washed twice with N2B27 supplemented with 10  $\mu\text{M}$  Y-27632. Flow cytometry was performed on the SONY MA900 cell sorter and data were analysed using FlowJo (v10.10.0).

## RNA extraction and RT–qPCR

RNA extraction was performed using TRIzol (15596018, Life Technologies) directly on dissociated hnPSCs carrying the indicated perturbation

or, in the case of blastoids and dark spheres, before RNA extraction, the structures were dissociated in a 1:1 mixture of trypsin-EDTA 0.5% (15-400-054, Fisher Scientific) and Accutase (07920, StemCell Technologies) for 5 min, diluted in N2B27 and spun down. Extraction was followed by RNA purification using a Direct-zol RNA-prep kit (R2052, Zymo Research) with DNase treatment. Of RNA, 1  $\mu\text{g}$  was retrotranscribed into cDNA using a SensiFAST cDNA synthesis kit (BIO-65053, Bioline), cDNA was diluted 1:4 with molecular grade water and 2  $\mu\text{l}$  of this dilution was used for quantitative PCR (qPCR) with primers for each amplicon (Supplementary Table 1). qPCR was performed in a LightCycler 480 Instrument (II) using a SensiFAST SYBR (Bioline, BIO-98020). For experiments using Taqman probes, qPCR Primetime probes were purchased from IDT (sequences in Supplementary Table 1) and were combined for qPCR with the LightCycler 480 Probes Master mix (04707494001, Roche).

## CUT&RUN

Protocol was performed according to Meers et al.<sup>69</sup> and using the CUTANA reagents from EpiCypher (concanavalin A-conjugated paramagnetic beads (21-1401), pAG-Tn5 (15-1017) and *Escherichia coli* spike-in DNA (18-1401)). We used 500,000 nontarg-CARGO or LTR5Hs-CARGO hnPSCs per condition. Cells were permeabilized using 0.005% of digitonin, and 0.5  $\mu\text{g}$  of H3K9me3 primary antibody was used per sample (Supplementary Table 1). DNA was extracted using phenol–chloroform, and library preparation was performed using the NEB-Next Ultra II Library Prep kit (E7645S, New England Biolabs). Libraries were sequenced paired-end for 150 cycles in a Novaseq 6000 Illumina sequencer.

## ChIP–seq, ChIP–seq libraries construction and sequencing

Cells were grown on feeder-free conditions using Geltrex to minimize MEF contamination in the sequencing. One  $10 \text{ cm}^2$  (approximately  $6 \times 10^6$  hnPSCs) was used per ChIP. Cells were crosslinked in PBS containing 1% methanol-free formaldehyde (28908, Pierce) for 10 min. Fixation was quenched during 10 min by adding a final concentration of 0.1 M of glycine. Upon harvesting, cells were resuspended in buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 and 0.25% Triton X-100) and incubated for 10 min, rotating at 4 °, before centrifugation at 1,350g for 5 min at 4 °C. The pellet was lysed in buffer 2 (10 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA and 0.5 mM EGTA), incubated for 10 min at 4 °C and once again centrifuged at 1,350g for 5 min. Then, the pellet was lysed in buffer 3 (10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate and 0.5% N-lauroylsarcosine), incubated for 20 min on ice and sonicated in a Bioruptor sonicator (Diagenode) until the obtention of DNA fragments of sizes ranging 400–600 bp. Chromatin was quantified, and approximately 10–25  $\mu\text{g}$  of chromatin were used for immunoprecipitation in a total of 500  $\mu\text{l}$  of buffer 3 containing the antibodies indicated in Supplementary Table 1. After overnight incubation, 100  $\mu\text{l}$  of magnetic protein G beads (10004D, Life Technologies) were added to each immunoprecipitation. After 2–3 h of incubation, the immunocomplexes were washed five times with RIPA wash buffer (50 mM HEPES-KOH pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% NP-40 and 0.7% sodium deoxycholate) and once with TE-NaCl buffer (50 mM Tris pH 8, 10 mM EDTA and 50 mM NaCl). To recover the DNA, the immunocomplexes were eluted in elution buffer (50 mM Tris pH 8, 10 mM EDTA and 1% SDS) at 65 °C for 15 min with vortexing every 5 min. The bead eluate was decrosslinked overnight at 65 °C. After RNase A treatment for 30 min (FEREN0531, Thermo Fisher Scientific) and proteinase K treatment (EO0492, Thermo Fisher) for 2 h, the DNA was purified using a Qiagen kit (28106, Qiagen).

To prepare ChIP–seq libraries for sequencing, we utilized the NEB-Next Ultra II DNA kit (E7645S, NEB), and Agencourt AMPure XP beads (A63881, Beckman Coulter) were used for the cleanings. We started from approximately 20 to 50 ng of ChIP or input DNA and followed the manufacturer’s instructions. Paired-end sequencing (150 cycles)

was performed in a Novaseq X Plus sequencer (Illumina) including 1% of PhiX.

### Bulk RNA-seq and library preparation

RNA was extracted using TRIzol from nontarg-CARGO and LTR5Hs-CARGO hnPSCs treated with cumate during 4 days in the absence of doxycycline, from ZNF729-FH hnPSCs treated with dTAG<sup>-1</sup> for 3 h and 24 h or from blastoids and dark spheres. mRNA was purified using poly-T oligo-attached magnetic beads. After fragmentation, the first-strand cDNA was synthesized using random hexamer primers followed by the second-strand cDNA synthesis. Libraries were prepared by end repair, A-tailing, adapter ligation, size selection, amplification and purification, and they were checked with Qubit and qPCR for quantification and Bioanalyzer for size-distribution detection. Quantified libraries will be pooled and sequenced on a Novaseq 6000 Illumina sequencer.

### Blastoid immunostainings

Immunostaining of blastoids was performed 'in well'. Media from Aggrewell were carefully aspirated (more than 90%). For fixation, 1 ml of 4% paraformaldehyde was added to the well and incubated at room temperature for 15 min. The paraformaldehyde was carefully aspirated and substituted for a rinse buffer composed of PBS with 3 mg ml<sup>-1</sup> polyvinylpyrrolidone. After one rinse, blastoids were permeabilized in PBS-polyvinylpyrrolidone containing 0.25% of Triton X-100 for 30 min. The permeabilization solution was aspirated substituted with blocking buffer (0.1% BSA (A9418, Sigma-Aldrich), 0.01% Tween 20 (P1379, Sigma-Aldrich) and 2% donkey serum (017-000-121, Jackson ImmunoResearch)), which was dispensed in the well with a 5-ml serological pipette to subsequently collect all the blastoids from the well and deposit them into a well of a six-well plate containing more blocking solution. Blocking took place for at least 3 h at 4 °C. Blastoids were picked using standard mouth pipetting or 20- $\mu$ l pipette tips and moved to primary antibodies (Supplementary Table 1) diluted in blocking solution in Nunc MicroWell MiniTrays (12-565-154, Fisher Scientific) at 4 °C overnight. Blastoids were washed three times with blocking buffer and stained with Alexa Fluor secondary antibodies for 3 h, washed three times and imaged in blocking buffer using an 18-well microslide (81826, Ibidi) in an Inverted Zeiss LSM 780 confocal microscope.

### PIP-seq

PIP-seq<sup>70</sup> is an alternative to microfluidics-based scRNA-seq methods that captures cells via vortex and can be performed from beginning to library preparation at the experimenter's bench. Blastoids from two wells of an Aggrewell plate per condition were collected and centrifuged in a 15-ml tube for 2 min at 250g, then the supernatant was aspirated. The blastoids pellet was resuspended in collagenase IV (07909, StemCell Technologies) and incubated at 37 °C with mild agitation for 40 min. Blastoids were centrifuged again in N2B27 medium and the pellet was resuspended in 0.5% trypsin-EDTA (15-400-054, Fisher Scientific) and incubated for 10 min at 37 °C. Two further washes were performed with N2B27, and the dissociated cells were passed through a 40- $\mu$ m Flowmi cell strainer. Experiment only continued when viability was larger than 80%. Cells (40,000 or less) were counted, captured and used for completing the PIP-seq T20 3' Single Cell RNA Kit protocol (FBS-SCR-T20-4-V4.05, Fluent Biosciences) without changes and using 12 cycles of cDNA amplification. Libraries were prepared with the reagents in the kit and were sequenced in an Illumina Novaseq X instrument.

### Western blot

After SDS-PAGE electrophoresis, protein transfer was carried out on a multilayered cassette including a nitrocellulose membrane. The transfer buffer was composed of 25 mM Tris-HCl, 192 mM glycine, 0.05% SDS and 10% methanol. The power source was set to 125 V for

90 min. The nitrocellulose membrane was blocked with 5% milk for 1 h and incubated with a HA tag antibody (Supplementary Table 1) overnight to detect ZNF729-HA or ZNF729-FH. For gel source data, see Supplementary Fig. 1.

### Image obtention and quantification

All bright-field images were taken using the EVOS FL Imaging System. The fluorescent immunostainings were imaged using an Inverted Zeiss LSM 780 confocal microscope. To obtain blastoids inner cell mass (ICM):trophectoderm (TE) ratios, we used the Fiji software<sup>71</sup>. We calculated the diameter of the blastoids cavity and the ICM size by measuring the distance from the point of contact with the trophectoderm to the end of the ICM. To count number of cells expressing specific lineage markers, we used a combination of software and manual counting. KLF17, GATA4 and cleaved CASP3-positive cells were counted with Fiji's cell counter in each stack. GATA3-positive cells were counted using the 3D Object Counter plug-in from Fiji, carefully curating the assigned positive cells with the human eye and correcting when necessary (for example, fluorescent artefacts that are not cells).

### Quantification of blastoid formation efficiency

For determining blastoid efficiencies, end-point (96 h) blastoids were moved to a 15-ml conical tube and the total volume was measured. Next, two technical duplicates of 50- $\mu$ l aliquots were dispensed into a 96-well plate and the structures were evaluated and counted, ultimately extrapolating to the total conical tube volume and to the 1,200 microwells present in the Aggrewell. To consider a 3D structure as a blastoid, we followed previously established criteria<sup>17</sup>. In brief, its morphology should resemble stage B6 of human blastocyst, with an accumulation of cells surrounded by a monolayered cyst mimicking the ICM and the trophectoderm, respectively. The ICM of the blastoids is often outside the cyst, in such case, we still considered that structure a blastoid. Blastoids have an approximate total diameter between 150  $\mu$ m and 250  $\mu$ m, and in the case of LTR5Hs-CARGO blastoids, the cavity should be larger than 150  $\mu$ m, with no upper limit. When tested by scRNA-seq or immunofluorescence, blastoids must express markers consistent with the lineages of blastocyst. Dark spheres are structures that appear darker in bright field and are not cavitated. We note that the efficiencies calculated may be an underestimation, as some aggregates or blastoids are accidentally aspirated during the medium changes.

### Recording of blastoid formation video

The blastoid formation protocol was performed as indicated above, with changes. Instead of using Aggrewells (which have an opaque bottom), we utilized Elplasia 24-well plates that allow imaging from below (4441, Corning). We note that the initial cell aggregation in these plates is not as robust, thus end-point blastoid formation is less efficient than in Aggrewells. Cell aggregates are monitored, and when there are early signs of cavitation (small 'bubbles' around the aggregates), the plate is moved to a Nikon Eclipse Ti-E microscope that is equipped with a system for CO<sub>2</sub> and temperature control (OKOLab). Blastoids were imaged at 37 °C and 5% CO<sub>2</sub> for 24 h.

### CUT&RUN analysis

Standard Illumina adapters were cut from the Illumina reads using Cutadapt<sup>72</sup> and then aligned to a combined hg38 and *E. coli* genome version using Bowtie2 (ref. 73), with the -dovetail parameter on and the rest of parameters in its default behaviour. This means that, in case of multimapping, all the valid alignments are reported. PCR duplicates were removed from the analysis. Coverage bigwig files were generated with Deeptools<sup>74</sup> bamCoverage and the -scaleFactor was set to the number obtained from the normalization of fragments mapped to the human genome (hg38) and the mapped fragments to the *E. coli* k12 MG1655 genome. Browser captures were obtained from IGV<sup>75</sup>.

## ChIP-seq analysis

Standard Illumina adapters were cut from the Illumina reads using Cutadapt<sup>72</sup>. Reads were aligned to the *Homo sapiens* (hg38) genome using Bowtie2 (ref. 73) in its default behaviour. PCR duplicates were removed from the analysis using Samtools<sup>76</sup>. Coverage bigwig files were generated with Deeptools<sup>74</sup> bamCoverage. Browser captures were obtained from IGV<sup>75</sup>.

Peaks were called using MACS3 (ref. 77). Identification of ZNF729-FH-bound repetitive DNA was performed by intersecting ZNF729-FH peaks with RepeatMasker (RRID:SCR\_012954)<sup>78</sup> using Bedtools<sup>79</sup> intersect with -f0.3. To be considered a peak at the promoter, ZNF729-FH or TRIM28 must bind to -1 kb or +200 bp around the transcription start site. Motif discovery analysis at the non-repetitive ZNF729 peaks was performed using the top 3,000 ZNF729-bound peaks using SeqPos<sup>58</sup>. Full SeqPos output can be found in Supplementary Table 7.

## RNA-seq and Gene Ontology analysis

Illumina adapters were trimmed from reads using Skewer<sup>80</sup>. Transcript alignment and quantification were performed using Salmon<sup>81</sup> against the human genome assembly version Gencode (v47)<sup>82</sup>. For differential gene expression analysis, we used DESeq2 (ref. 83) after excluding transcripts with less than 10 reads across the tested samples. DESeq2 compared the effect of nontarg-CARGO and LTR5Hs-CARGO in hnPSCs, the differences between blastoids and dark spheres, or the gene expression changes upon dTAG<sup>-1</sup> addition to the ZNF729-FH hnPSCs. Biological replicates were used as covariates. Analysis of chimpanzee naive PSCs bulk RNA-seq<sup>39</sup> was performed in the same manner but using the *Pan troglodytes* panTro6 Clint\_PTRv2 genome assembly. The rhesus genome Mmul\_10 (RheMac10) genome reference lacks a *ZNF729* transcript model. To assess the presence and expression of *ZNF729* in the rhesus macaques naive state, we performed unguided transcriptome assembly from rhesus monkey (*Macaca mulatta*) naive PSCs bulk RNA-seq data<sup>40</sup> using the Trinity pipeline<sup>84</sup>. We constructed a blast database from the Trinity output and searched for the human *ZNF729* nucleotide sequence using blastn and blastx algorithms<sup>85</sup>. The highest matches were searched against non-redundant NCBI database with blastn and blastx algorithms. None of the sequences scored *ZNF729* in reciprocal blast as a top match. When a sequence had a blast matching to *ZNF729*, such matches were below 60% identity. Differentially regulated genes (FDR 5%) were used for human Gene Ontology analysis using Gorilla<sup>86</sup> using as a background the list of genes expressed in hnPSCs, blastoids and dark spheres.

## TEtranscripts

The software 'TEtranscripts'<sup>87</sup> was used to find differentially regulated transposable elements in the ZNF729-FH dTAG<sup>-1</sup> bulk RNA-seq experiments. To this end, the RNA-seq reads were aligned using HISAT2 (ref. 88), and following the tool's manual, we allowed 100 alignments per read (-k 100) to optimize transposable element quantification and differential analysis. This analysis searches for differences at the level of transposable element families, so we cannot exclude the possibility that within a family, specific individual insertions could still be affected by ZNF729 depletion.

## PIP-seq and pseudobulk differential gene expression analysis

For each sample and replicate, reads obtained from the Novaseq X were analysed with Fluent Bio's proprietary software Pipseeker with default parameters and aligning against the GRCh38 transcriptome index (Gencode v40 2022.04, Ensembl106). A background removal step was performed in all samples using CellBender<sup>89</sup> with parameters --fpr 0.01 and --epochs 150. Full count matrices with background RNA removed were further analysed using Seurat<sup>90</sup>. Cells with more than 10% of mitochondrial counts were eliminated and the number of genes detected was also used for filtering the data (see Supplementary Table 3 for

specific parameters applied for each sample). Each object was normalized using Seurat's LogNormalize method and transformed using the ScaleData function removing the unwanted variation originated from mitochondrial contamination or the cell-cycle stage. Upon examining elbow plots, 20 principal components were considered significant for unsupervised clustering with the FindNeighbors and FindClusters Seurat functions. Cluster identities were assigned based on the genes specifically marking each cluster according to Seurat's FindMarkers function and comparing them with lineage markers uncovered in human pre-implantation datasets<sup>18,27</sup>. Seurat objects belonging to the nontarg-CARGO and the LTR5Hs-CARGO blastoids were merged and the LTR5Hs-CARGO object was downsampled to the same number of cells as the nontarg-CARGO object for comparison purposes. Multiple iterations of downsampling were performed with comparable results. We subset cells belonging to the epiblast or the neo-epiblast into a single group and performed differential gene expression analysis using DESeq2 on the sample-level aggregated counts (pseudobulk). Specifically, DESeq2 tested the effect of the repression of LTR5Hs elements (nontarg-CARGO versus LTR5Hs-CARGO) using the PIP-seq replicate as a covariate. Only genes with FDR of 5% and fold changes < -1 or fold changes > 1 were considered statistically significant. Genes related to the trophoctoderm or placental development in Extended Data Fig. 5g are established markers or were obtained from Petropoulos et al.<sup>18</sup> and other literature searches. Regarding the number of cells belonging to the different clusters, we do note that scRNA-seq may not be accurate for trophoctoderm cell counting, as we have observed an accelerated lysis of trophoctoderm cells upon blastoid dissociation before cell capture. In agreement with such possibility, the overall proportion of trophoctoderm cells was systematically lower in our scRNA-seq analyses than in immunostainings, irrespective of the LTR5Hs activity status.

## Projection of transcriptomes into human embryo datasets

To identify the human embryo counterparts of the transcriptomes of cells dissociated from nontarg-CARGO or LTR5Hs-CARGO blastoids, we projected such transcriptomes into a collection of human embryo reference datasets<sup>18,27,91-94</sup>. Counts from each gene in each cell (slot 'counts' in Seurat) were extracted and uploaded to a human embryogenesis online prediction tool<sup>51</sup> (<https://petropoulos-lanner-labs.clintec.ki.se/>). The identified annotations and UMAP values were used in our plots and conclusions.

## scRNA-seq data analysis including transposons

Raw data from Kagawa et al.<sup>17</sup> was downloaded from the Gene Expression Omnibus database entry GSE177689. Smart-seq2 PCR adapters were trimmed using Skewer<sup>80</sup> and the resulting reads were aligned using HISAT2 (ref. 88) with the parameters --dta --no-mixed --no-discordant -k 100, to allow enough multimappers for transposon analysis. The resulting bam files were processed with the scTE<sup>95</sup> software for transposon family identification and quantification. The resulting matrix was subset to contain only cells dissociated from the 96-h blastoids and it was analysed using Seurat, filtering cells with more than 25% of mitochondrial counts, less than 2,000 or more than 16,000 genes. Downstream unsupervised clustering was performed using 20 principal components. For comparing blastoids from Kagawa et al. and this article, we integrated a nontarg-CARGO scRNA-seq object with the data from Kagawa et al. after removing the transposons using default Seurat data integration functions.

## Human-mouse and human-marmoset comparisons

LTR5Hs-regulated genes in the blastoids epiblast located within 250 kb of an LTR5Hs element were used as LTR5Hs target genes. To obtain genes expressed in the mouse and the marmoset epiblast, we used a previously published table containing expression levels and orthology analysis of human, mouse and marmoset genes. Specifically, we



used the epiblast data (early ICM in the dataset)<sup>34</sup>. Only genes with an average expression of more than 2 fragments per kilobase million were considered expressed, with a caveat that this analysis may pass over more subtle, quantitative differences in expression between species, which can nonetheless be functionally important. This cut-off was validated by visual inspection of scRNA-seq datasets from mouse embryos<sup>96</sup>. Genes were assigned to evolutionary branches using data from the Gentree database<sup>38</sup>.

### ZNF729 locus conservation

Figure 3e is a cartoon depiction of Zoonomia project's Cactus genomic alignments<sup>53,54</sup> in the UCSC genome browser<sup>97</sup>.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Datasets generated in this article have been deposited in the Gene Expression Omnibus repository: GSE262191 for CUT&RUN, GSE296554 for bulk RNA-seq, GSE262329 for PIP-seq and GSE296555 for ChIP-seq. Source data are provided with this paper.

### Code availability

No custom code was used; only previously published pipelines were implemented.

59. Lovell-Badge, R. et al. ISSCR guidelines for stem cell research and clinical translation: the 2021 update. *Stem Cell Rep.* **16**, 1398–1408 (2021).
60. Masaki, H. et al. Interspecific in vitro assay for the chimera-forming ability of human pluripotent stem cells. *Development* **142**, 3222–3230 (2015).
61. Mullick, A. et al. The cumate gene-switch: a system for regulated expression in mammalian cells. *BMC Biotechnol.* **6**, 43 (2006).
62. Nabet, B. et al. The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
63. Nabet, B. et al. Rapid and direct control of target protein levels with VHL-recruiting dTAG molecules. *Nat. Commun.* **11**, 4687 (2020).
64. Martin, R. M. et al. Highly efficient and marker-free genome editing of human pluripotent stem cells by CRISPR-Cas9 RNP and AAV6 nonor-mediated homologous recombination. *Cell Stem Cell* **24**, 821–828.e5. (2019).
65. Gregorevic, P. et al. Systemic delivery of genes to striated muscles using adeno-associated viral vectors. *Nat. Med.* **10**, 828–834 (2004).
66. Gray, J. T. & Zolotukhin, S. Design and construction of functional AAV vectors. *Methods Mol. Biol.* **807**, 25–46 (2011).
67. Kagawa, H. et al. Protocol for human blastoids modeling blastocyst development and implantation. *J. Vis. Exp.* <https://doi.org/10.3791/63388> (2022).
68. Io, S., Iemura, Y. & Takashima, Y. Optimized protocol for naive human pluripotent stem cell-derived trophoblast induction. *STAR Protoc.* **2**, 100921 (2021).
69. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314 (2019).
70. Clark, I. C. et al. Microfluidics-free single-cell genomics with templated emulsification. *Nat. Biotechnol.* **41**, 1557–1566 (2023).
71. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
72. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12 (2011).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
75. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
76. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. Zhang, Y. et al. Model-based analysis of Chip-seq (MACS). *Genome Biol.* **9**, R137 (2008).
78. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).

79. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
80. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
81. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417–419 (2017).
82. Mudge, J. M. et al. GENCODE 2025: reference gene annotation for human and mouse. *Nucleic Acids Res.* **53**, D966–D975 (2025).
83. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
84. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
85. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
86. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
87. Jin, Y., Tam, O. H., Paniagua, E. & Hammell, M. Tetrascripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 (2015).
88. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
89. Fleming, S. J. et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
90. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
91. Yanagida, A. et al. Naive stem cell blastocyst model captures human embryo lineage segregation. *Cell Stem Cell* **28**, 1016–1022.e4 (2021).
92. Tyser, R. C. V. et al. Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).
93. Xiang, L. et al. A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* **577**, 537–542 (2020).
94. Yan, L. et al. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
95. He, J. et al. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat. Commun.* **10**, 34 (2019).
96. Nowotschin, S. et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361–367 (2019).
97. Perez, G. et al. The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res.* **53**, D1243–D1249 (2025).
98. Torre, D. et al. Isoform-resolved transcriptome of the human preimplantation embryo. *Nat. Commun.* **14**, 6902 (2023).
99. Tribollet-Hardy, J. et al. Genetic features and genomic targets of human KRAB-zinc finger proteins. *Genome Res.* **33**, 1409–1423 (2023).

**Acknowledgements** We thank members of the Wysocka laboratory for their continued help and critical reading of the manuscript, especially to S. Tabatabaee for technical help with recording the blastoids video; F. Suchy and J. Bhadury for sharing technical expertise on human embryogenesis; H. Masaki for generating the hnPSCs; C. Estaras and L. Colella for helpful discussions; and A. Jolma and T. Hughes for sharing unpublished data on ZNF729. The cartoons in Figs. 1c, 3c and 5d and Extended Data Figs. 1a, 3d and 7a were created using BioRender (<https://biorender.com>). The HERVKcon plasmid was a gift from P. Bieniasz<sup>24</sup>, the pAAV-GFP plasmid was a gift from J. Gray (Addgene #32395) and the pDGM6 plasmid was a gift from D. Russell (Addgene #110660). R.F. was funded by an EMBO long-term postdoctoral fellowship (ALT-1-2019) and a Cancer Research Institute–Bristol Myers Squibb fellowship. S.W. was supported by a Stanford Graduate Fellowship. O.J.C. was funded by a National Science Foundation Graduate Research Fellowship Program. J.W. was funded by the Howard Hughes Medical Institute, a Lorry Lokey endowed professorship, Stanford Ludwig Center for Cancer Stem Cell Research and Medicine and the Stinehart Reed seed grant.

**Author contributions** R.F. and J.W. conceptualized the study. R.F. designed, executed and analysed all the experiments except for the flow cytometry measurements. S.W. helped engineer the hnPSCs and with the blastoid stainings. O.J.C. performed the flow cytometry experiments. T.S. contributed to the computational analyses. J.W. supervised the work. R.F. and T.S. performed the statistical analyses. H.N. reviewed the study. The manuscript was written by R.F. and J.W., with contributions from other co-authors.

**Competing interests** J.W. is a paid member of the Camp4 scientific advisory board. All other authors declare no competing interests.

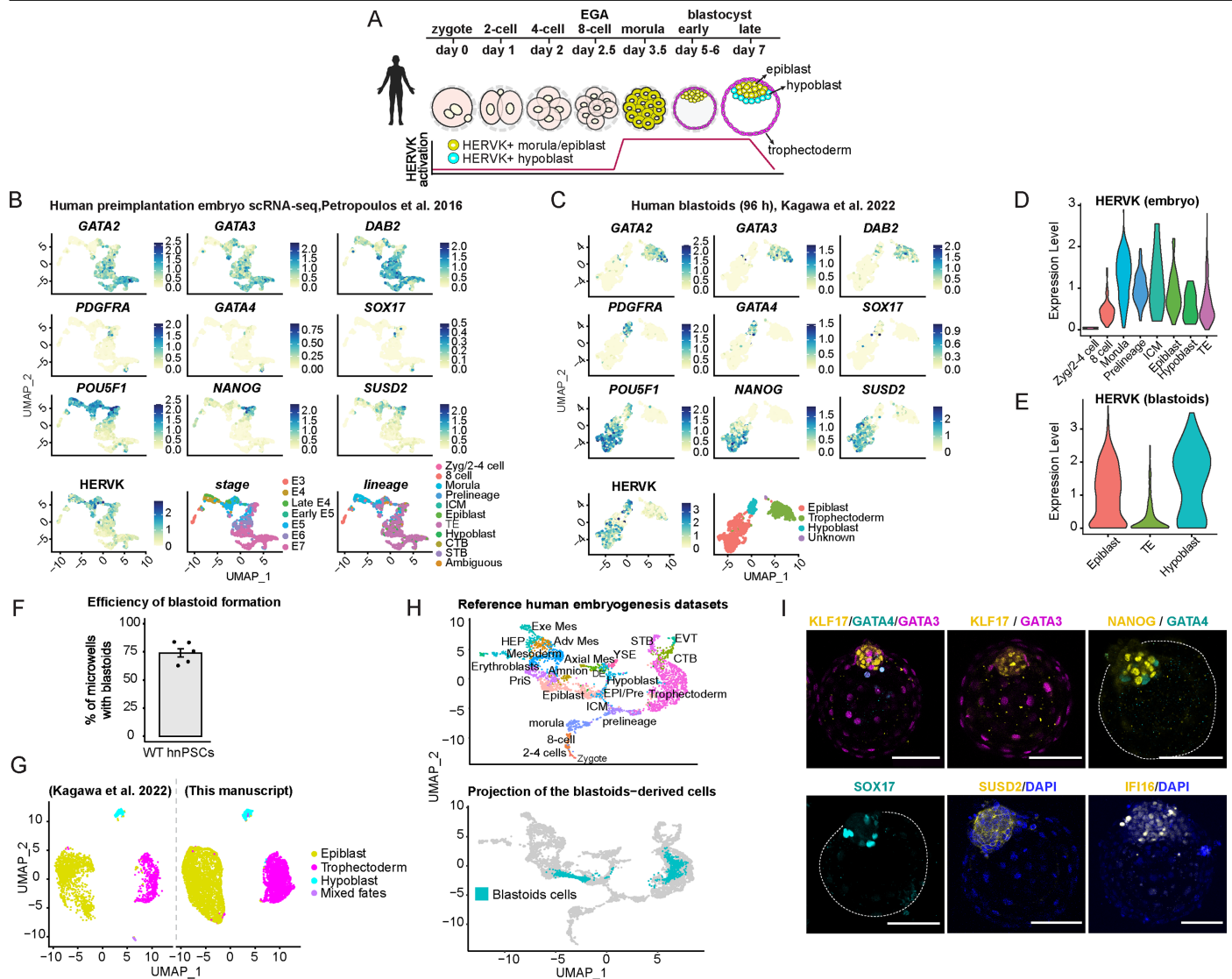
### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09571-1>.

**Correspondence and requests for materials** should be addressed to Joanna Wysocka.

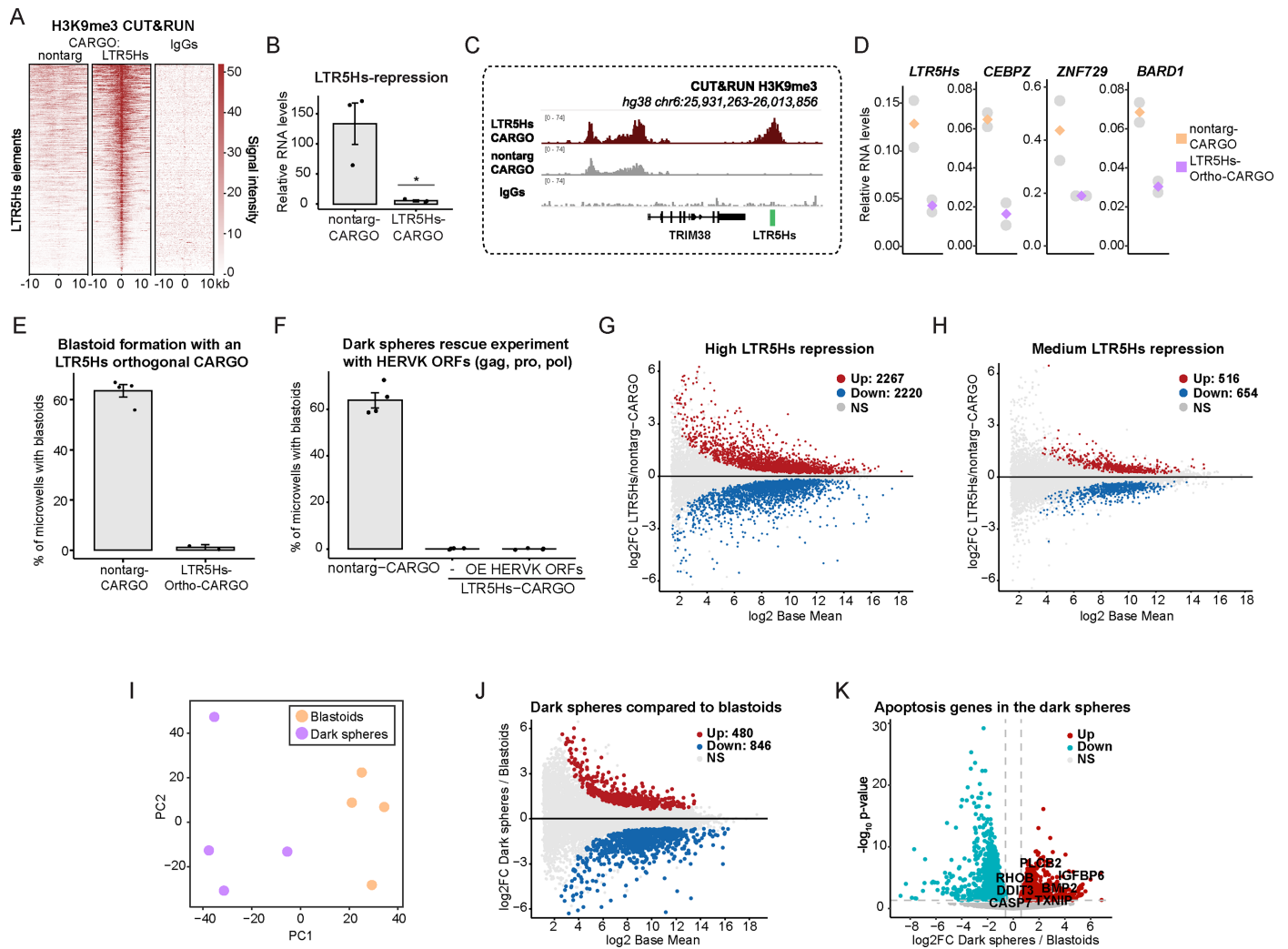
**Peer review information** Nature thanks Vincent Pasque, Tsutomu Sawai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



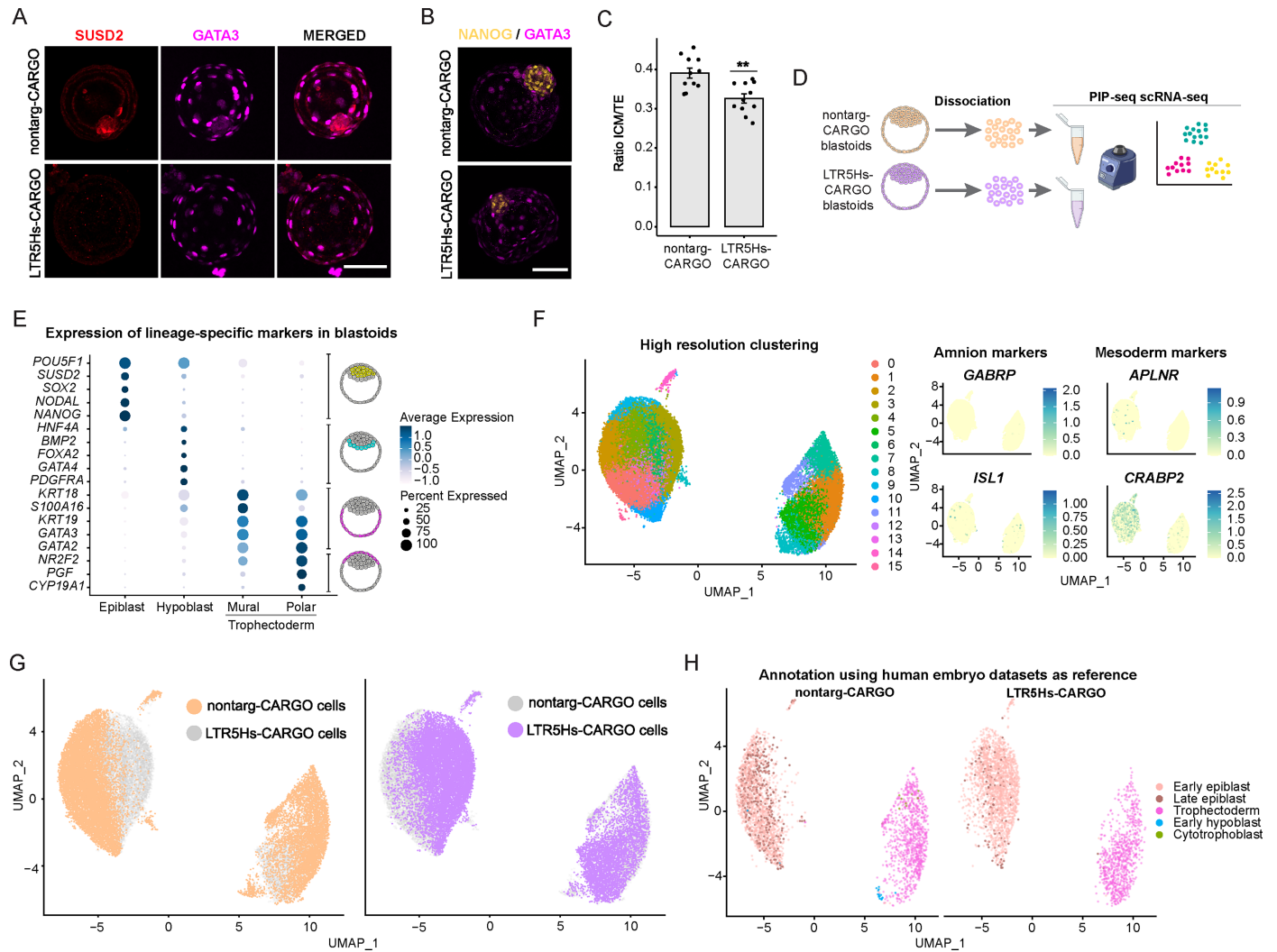
**Extended Data Fig. 1 | Benchmarking a human blastoid protocol to study HERVK in human pre-implantation development.** **a.** Stages of human pre-implantation development. HERVK expression levels from<sup>9,18</sup> are represented by a red line. HERVK is active in the epiblast (yellow cells) and the hypoblast (cyan cells). EGA: embryonic genome activation. The silhouette of the human was created in BioRender. Fueyo, R. (2025) <https://BioRender.com/tmgd0pn>. **b.** UMAP of expression of the indicated lineage markers and HERVK in human pre-implantation single cells. Bottom right two panels display annotation of data using stage (the prefix E is equivalent to days) or lineages (data and annotations from refs. 18, 51). **c.** UMAP of expression of the indicated lineage markers and HERVK in single cells dissociated from blastoids at 96 h. Bottom right panels display annotation of data using lineages (data and annotations from ref. 17). **d.** and **e.** Violin plots of HERVK normalized expression levels at the different

lineages present in human pre-implantation embryos<sup>18</sup> (**d**) or 96 h blastoids<sup>17</sup> (**e**). **f.** Bar plot displays efficiency of blastoid formation with wild-type hnPSCs,  $n = 5$  biologically independent replicates. Data represents mean values  $\pm$  SEM. **g.** UMAPs representing integrated single cell RNA-seq datasets from the previously reported human blastoid derivation protocol (Kagawa et al.<sup>17</sup>) and blastoids generated and benchmarked in this manuscript. Colors represent the blastoid lineages. **h.** UMAP of a reference collection of human embryo single cell RNA-seq transcriptomes<sup>51</sup> (top) and projection of cells from blastoids derived in this manuscript (bottom). **i.** Representative confocal images ( $n \geq 3$ ) of immunostaining of blastoids derived from wild-type hnPSCs with lineage-specific markers: KLF17, NANOG, SUSD2, IFI16 (yellow, epiblast), GATA4, SOX17 (cyan, hypoblast), and GATA3 (magenta, trophectoderm). White bar represents 100  $\mu$ m.



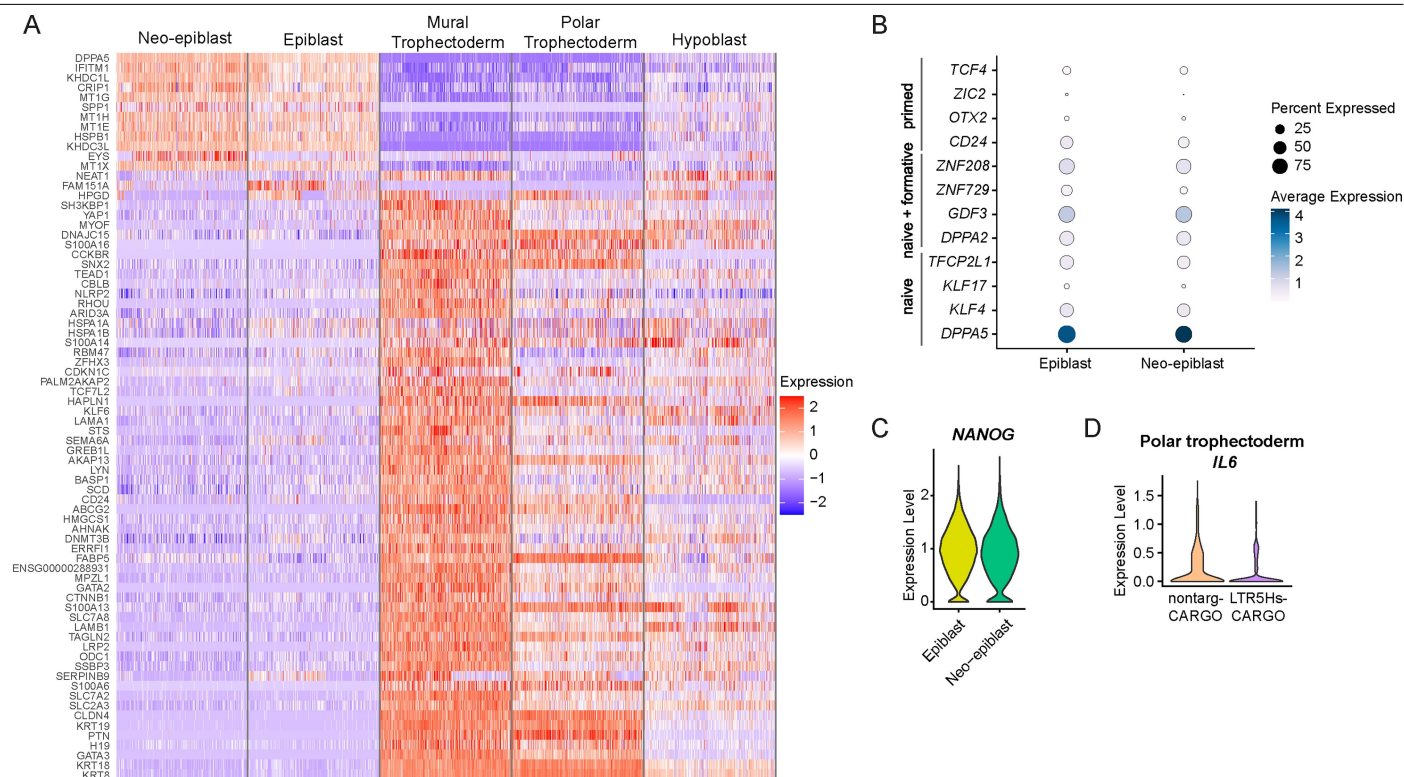
**Extended Data Fig. 2 | LTR5Hs-driven gene regulatory changes underlie blastoid formation potential.** **a.** Heatmaps displaying H3K9me3 CUT&RUN signal over LTR5Hs elements in nontarg-, LTR5Hs-CARGO cells, and in the IgGs negative control. **b.** RT-qPCR depicting LTR5Hs RNA levels measured with Taqman and normalized to the *RPL13* gene in nontarg-CARGO and LTR5Hs-CARGO cells,  $n = 3$ , unpaired two-tailed Student's  $t$ -test,  $P = 0.02$ . Data represents mean values  $\pm$  SEM. **c.** IGV genome browser capture of H3K9me3 CUT&RUN profiles over an LTR5Hs insertion. **d.** Plots depict qPCR RNA levels of LTR5Hs, and LTR5Hs-regulated genes in nontarg-CARGO and LTR5Hs-Ortho-CARGO hnPSCs, ( $n = 2$ ). **e.** Blastoid formation efficiencies of nontarg-CARGO and LTR5Hs-Ortho-CARGO hnPSCs ( $n = 4$  biological replicates for nontarg-CARGO and  $n = 2$  for LTR5Hs-Ortho CARGO). Data represents mean values  $\pm$  SEM. **f.** Blastoid formation efficiencies of two nontarg-CARGO hnPSCs clones, two LTR5Hs-CARGO high repression clones, and two LTR5Hs-CARGO high repression

clones expressing a transgene encoding the HERVK proteins gag, pro, and pol ( $n = 2$  biological replicates). Data represents mean values  $\pm$  SEM. **g.** MA plot representing bulk RNA-seq results of LTR5Hs-CARGO high repression clones ( $n = 4$  clones, in biological triplicates) compared to nontarg-CARGO clones ( $n = 3$ ). FDR 5%. **h.** MA plot representing bulk RNA-seq results of LTR5Hs-CARGO medium repression clones ( $n = 5$ , in biological duplicates or triplicates) compared to nontarg-CARGO clones. FDR 5%. **i.** PCA of the bulk RNA-seq transcriptomes obtained from blastoids and dark spheres upon LTR5Hs-CARGO induction for 96 h ( $n = 4$ ). **j.** MA plot representing bulk RNA-seq results of blastoids and dark spheres obtained upon LTR5Hs repression ( $n = 4$ ) FDR 5%. **k.** Volcano plot of bulk RNA-seq of dark spheres compared to blastoids ( $n = 4$ ). Labeled dots represent statistically significant genes involved in apoptosis according to GSEA curated gene sets<sup>25</sup>. DESeq2 Wald test FDR 5%. P-values are corrected for multiple testing using the Benjamini and Hochberg method<sup>52</sup>.



**Extended Data Fig. 3 | LTR5Hs-repression in human blastoids impairs lineage identity.** **a.** Representative confocal images of a nontarg-CARGO and LTR5Hs-CARGO blastoid immunostained with antibodies against the epiblast marker SUSD2 (red) and the trophoblast marker GATA3 (magenta),  $n = 3$  blastoids from two independent biological replicates. White bar represents 100  $\mu\text{m}$ . **b.** Representative confocal images of a nontarg-CARGO and LTR5Hs-CARGO blastoid immunostained with antibodies against the epiblast marker NANOG (yellow) and the trophoblast marker GATA3 (magenta)  $n = 3$ , white bar represents 100  $\mu\text{m}$ . **c.** Bar plots representing the ratio between the size of the inner cell mass (ICM) and the size of the trophoblast cavity (TE) in nontarg-CARGO and LTR5Hs-CARGO blastoids. Each dot represents measures from tens of blastoids in each independent biological replicate ( $n = 11$ , 7 biological replicates but 4 of them with technical replicates). Unpaired two-tailed Student's  $t$ -test,  $P = 0.0012$ . Data represents mean values  $\pm$  SEM. **d.** Schematic of

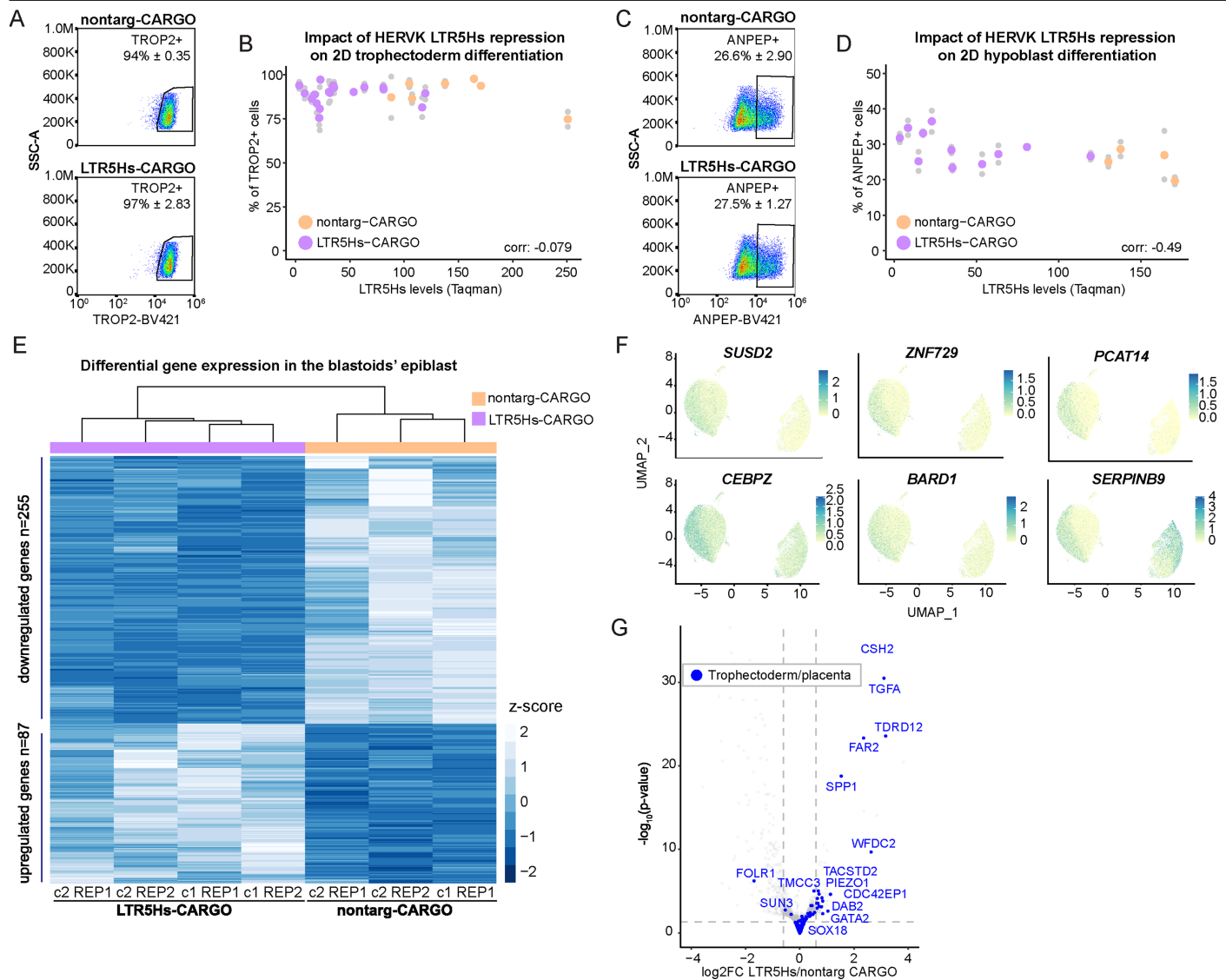
particle-templated instant partition sequencing (PIP-seq) of single cells obtained from nontarg- or LTR5Hs-CARGO blastoid dissociation. The PIP-seq illustrations were created in BioRender. Fueyo, R. (2025) <https://BioRender.com/3g5yuqw>. **e.** Dot plot showing the expression of genes indicated on the y-axis in cells assigned to the clusters indicated on the x-axis. Cartoons on the right highlight the location of cells expressing the subset of genes in the human blastocysts/blastoids. **f.** High resolution clustering (resolution = 1) of PIP-seq results (left) and UMAP of the expression of amnion and mesoderm markers (right) in those clusters. **g.** UMAP of transcriptomes from single cells dissociated from blastoids and colored by their origin for better visualization of the UMAP depicted in Fig. 2d, nontarg-CARGO blastoids (left, orange) and LTR5Hs-CARGO (right, purple). **h.** UMAP representing the nontarg-CARGO and LTR5Hs-CARGO transcriptomes colored by the assigned human embryo transcriptomic counterpart using the computational tool published in ref. 51.



**Extended Data Fig. 4 | Lineage identity features of LTR5Hs-repressed blastoids.** **a.** Heatmap representing the expression of genes identified as markers of the clusters indicated on top in the PIP-seq data. Color scale represents z-score. **b.** Dot plot showing the expression of naive, formative and primed pluripotency genes indicated on the y-axis in cells assigned to the epiblast or neo-epiblast

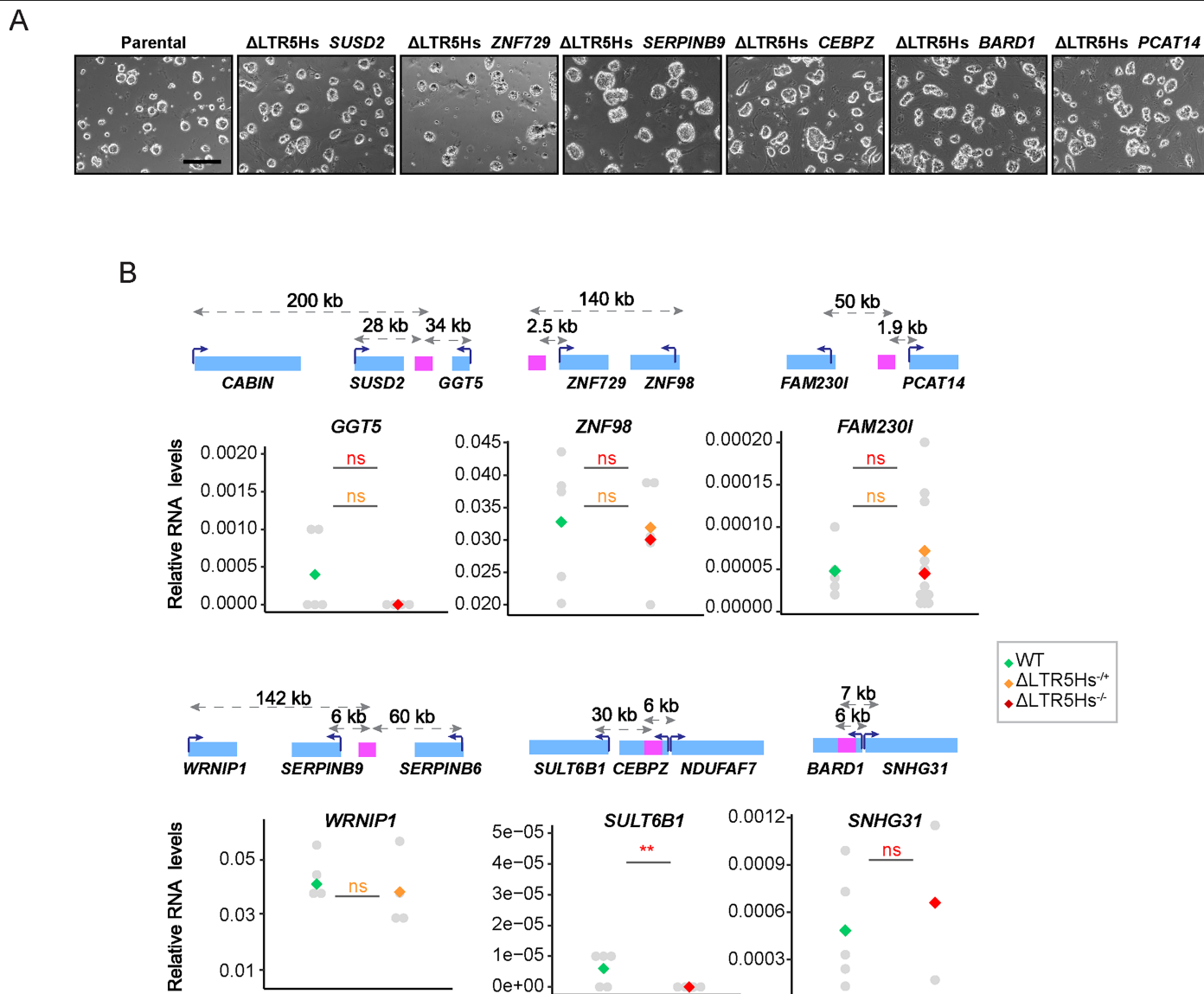
clusters indicated on the x-axis. **c.** Violin plot of *NANOG* normalized expression levels in the epiblast and neo-epiblast clusters. **d.** Violin plot of *IL6* normalized expression levels in the polar trophectoderm of nontarg-CARGO and LTR5Hs-CARGO blastoids.





**Extended Data Fig. 5 | LTR5Hs regulates gene expression in the epiblast of blastoids.** **a.** Flow cytometry density plots representing the analysis of TROP2+ cells compared to size scatter in 2D cultures of nontarg-CARGO (7 clonal cell lines in biological duplicates) and LTR5Hs-CARGO hnPSCs (15 clonal cell lines in biological duplicates) upon 3 days of differentiation using the trophectoderm protocol described in ref. 29. **b.** Representation of the quantitative results of the flow cytometry experiments described in **a.** Correlation represents Pearson's. **c.** Flow cytometry density plots representing the analysis of ANPEP+ cells compared to size scatter in 2D cultures of nontarg-CARGO (4 clonal cell lines in biological duplicates) and LTR5Hs-CARGO hnPSCs (11 clonal cell lines in biological duplicates) upon 3 days of differentiation using the hypoblast protocol described in ref. 55. **d.** Representation of the quantitative results of the flow cytometry experiments described in **c.** Correlation represents Pearson's. **e.** Heatmap and hierarchical clustering representing differentially expressed genes in the

blastoids generated from nontarg-CARGO or LTR5Hs-CARGO hnPSCs cell lines (two clonal cell lines from each condition, n = 2 PIP- seq biological replicates per clone, except for nontarg-CARGO clone 1, n = 1). Color scale represents z-score. **f.** UMAP of the selected downregulated genes showing higher expression in the blastoids epiblast (left cluster) compared to the neo-epiblast cluster (right). **g.** Volcano plot representing gene expression changes in LTR5Hs- versus nontarg-CARGO blastoid epiblast cells (using combined cells from epiblast and neo-epiblast clusters). Only genes from the trophectoderm or placenta lineages obtained from literature searches and refs. 18, 27 are colored in dark blue, the rest of genes are grey color. Vertical dashed lines indicate fold change < -1.5 or > 1.5 and horizontal dashed line indicates -log<sub>10</sub> of P = 0.05. DESeq2 Wald test FDR 5%. P-values are corrected for multiple testing using the Benjamini and Hochberg method<sup>52</sup>.

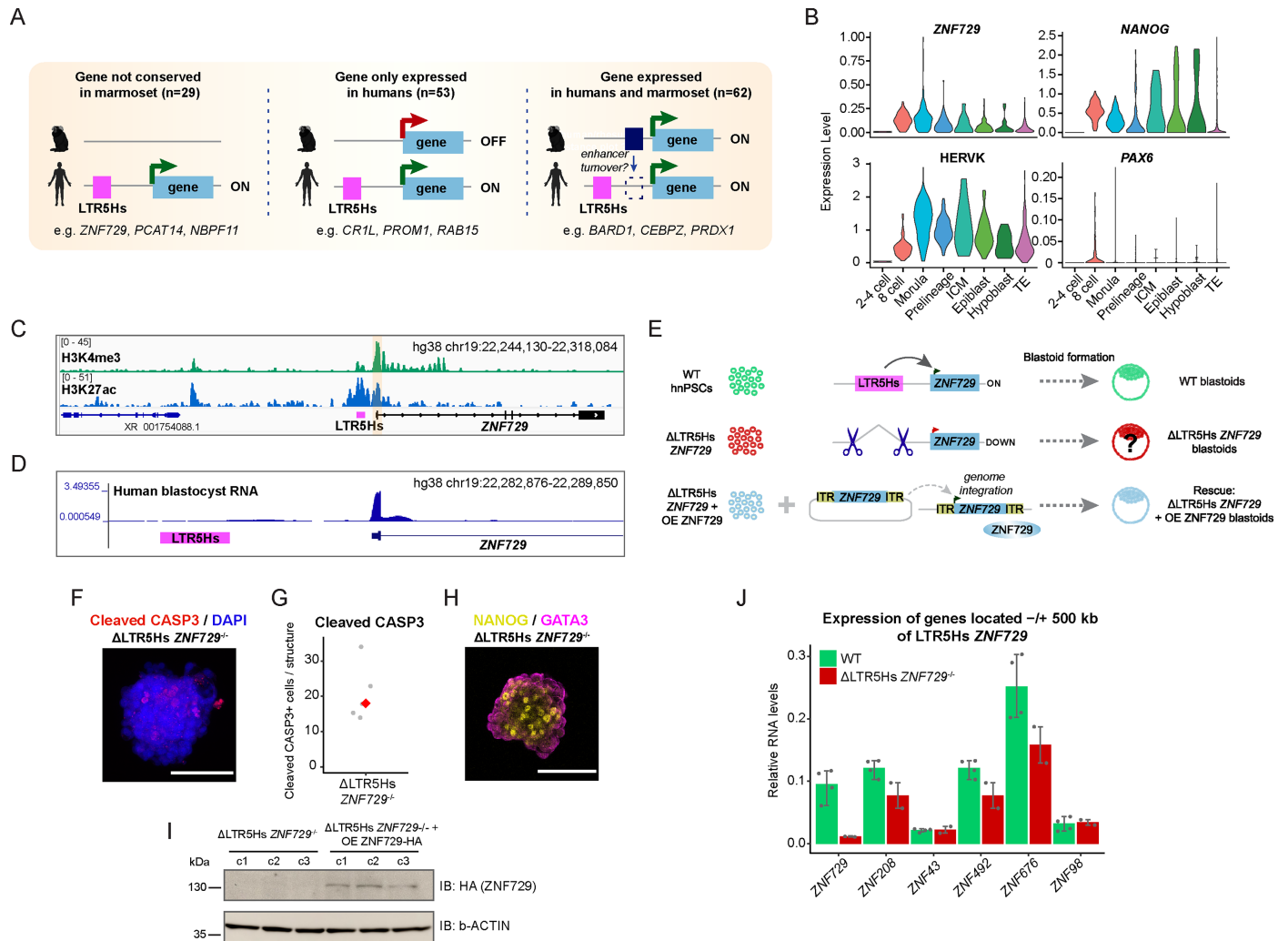


**Extended Data Fig. 6 | Expanded characterization of the  $\Delta$ LTR5Hs hnPSCs.**

**a.** Bright field images of  $\Delta$ LTR5Hs hnPSCs with the genes affected by each deletion indicated above each image. Black bar represents 200  $\mu$ m. These images are representative of the cells morphology in culture, n = 2 biological replicates.

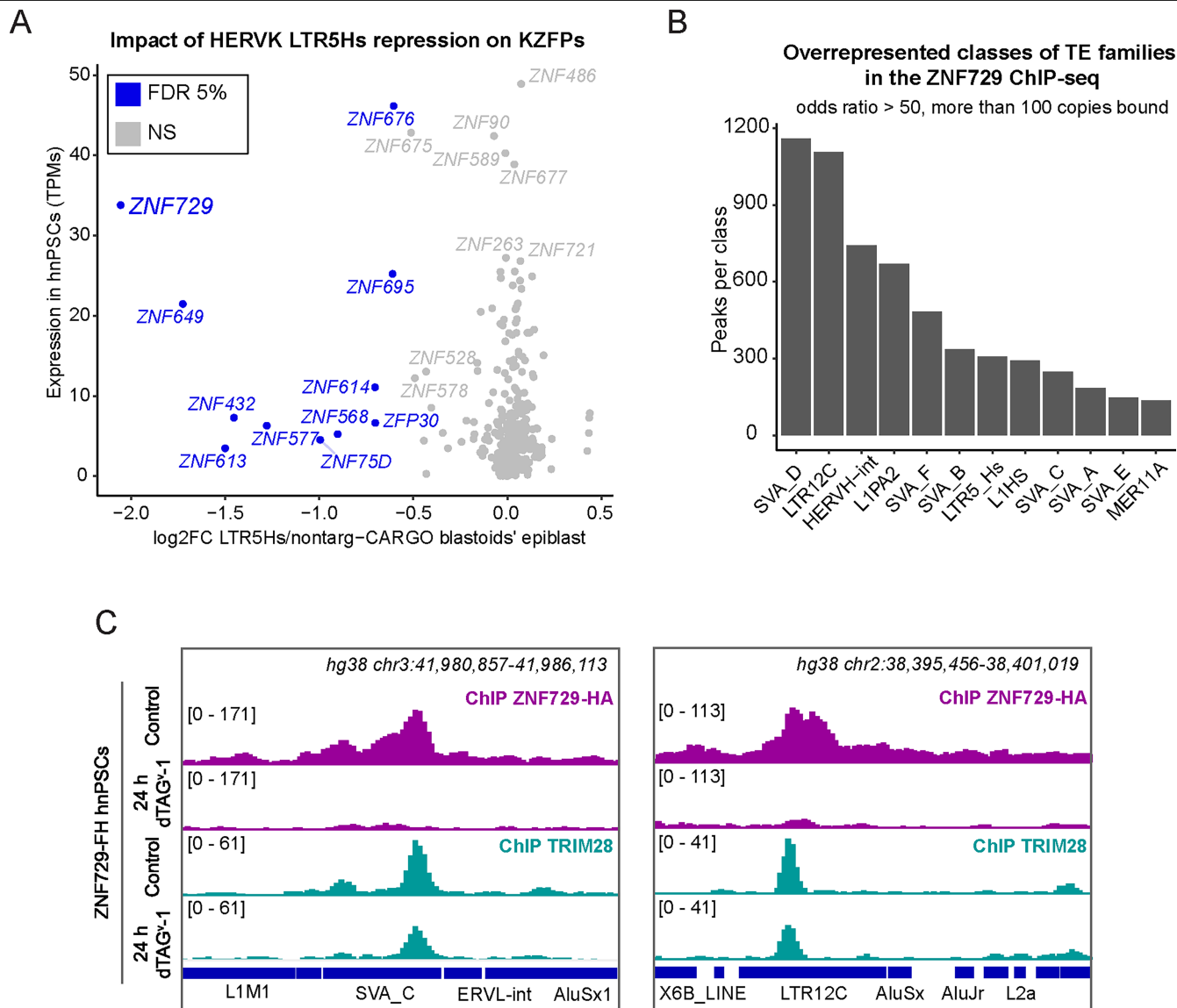
**b.** RT-qPCR results of genes unaffected by LTR5Hs deletions at gene loci shown in Fig. 3b. Plots display the expression of the indicated genes in wild-type or  $\Delta$ LTR5Hs hnPSCs. RNA values are normalized relative to *RPL13A*. Above each plot, schematic depiction of the locus is shown. Blue rectangles indicate genes, pink indicates the closest LTR5Hs element, the number on top of the dashed arrows

displays distance from the promoter to the LTR5Hs. Grey dots represent expression values obtained in each clone, green, orange, and red squares represent median values. N (independent clones) per gene: *GGT5*: WT = 5, heterozygous=2, homozygous=5; *ZNF98*: WT = 5, heterozygous=4, homozygous=2; *WRNIP1*: WT = 5, heterozygous=4; *SULT6B1*: WT = 5, heterozygous=4; *SNHG31*: WT = 5, heterozygous=2; *FAM230I*: WT = 5, heterozygous=6, homozygous=6. Unpaired two-tailed Student's t-test, asterisk color indicates homozygous (red) or heterozygous (yellow), specific p-values included as Source Data, ns: not significant.



**Extended Data Fig. 7 | Diversification of the human epiblast transcriptome by LTR5Hs and characterization of the  $\Delta$ LTR5Hs ZNF729 hnPSCs.** **a.** Epiblast expression and evolutionary conservation of LTR5Hs-regulated genes in marmosets, data published in ref. 34. The silhouettes of the human and marmoset were created in BioRender. Fueyo, R. (2025) <https://BioRender.com/tmgd0pn>. **b.** ZNF729 and HERVK normalized expression levels in human pre-implantation scRNA-seq data<sup>18</sup>. NANOG and PAX6 are depicted as a reference for a highly or lowly expressed transcription factor, respectively. **c.** IGV browser capture showing ChIP-seq signal of H3K4me3 (top, green) and H3K27ac (bottom, blue) at the LTR5Hs ZNF729 locus. **d.** UCSC browser capture of isoform resolved transcriptomic data in human blastocysts<sup>98</sup> at the LTR5Hs ZNF729 locus. **e.** Approach to address the essentiality of the human-specific LTR5Hs ZNF729 in blastoid formation. Top section represents wild-type hnPSCs (green). Middle section shows  $\Delta$ LTR5Hs ZNF729 hnPSCs (red, i.e. LTR5Hs has been homozygously or heterozygously deleted). Bottom section represents the rescue experiment,

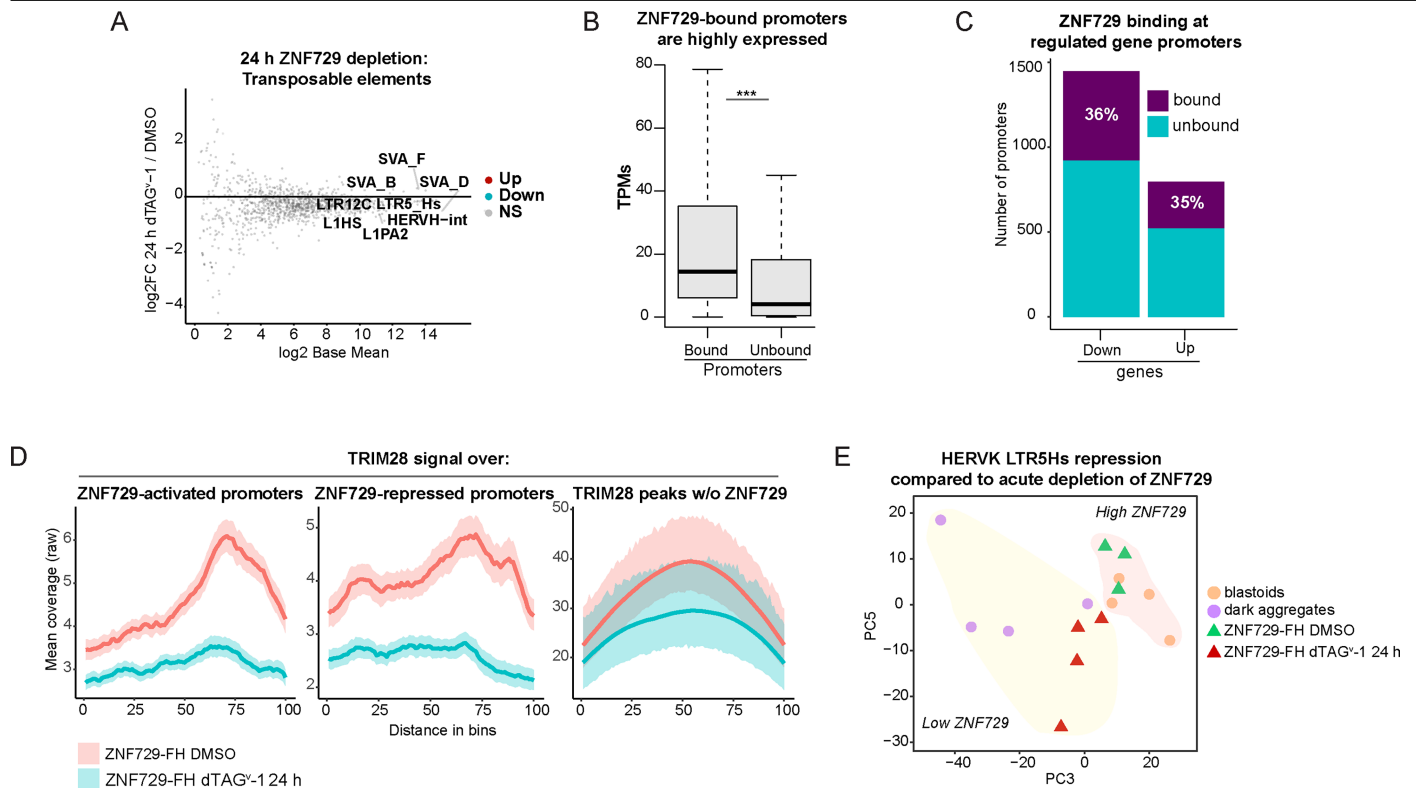
performed by overexpressing ZNF729 (blue). **f.** Representative  $\Delta$ LTR5Hs ZNF729<sup>-/-</sup> dark sphere immunostained with cleaved-CASP3 (red) and DAPI (blue). White bar represents 100  $\mu$ m. **g.** Quantification of cleaved-CASP3 immunostaining described in **f** ( $n = 5$  stained dark spheres, two biological replicates). **h.** Representative  $\Delta$ LTR5Hs ZNF729<sup>-/-</sup> dark sphere immunostained with NANOG and GATA3,  $n = 3$  biological replicates, white bar represents 100  $\mu$ m. **i.** Western blot validating the expression of the ZNF729 protein from the integrated ZNF729-HA cDNA transgene in three clonal cell lines of  $\Delta$ LTR5Hs ZNF729 hnPSCs. Membrane blotted with antibodies against the HA tag and  $\beta$ -ACTIN as loading control blotted on the same membrane.  $N = 2$  biological replicates. For gel source data, see Supplementary Fig. 1. **j.** RT-qPCR expression analysis of genes located within -1 Mb of the LTR5Hs ZNF729 locus. Plot displays the expression of the indicated genes in wild-type (green,  $n = 4$ ) or  $\Delta$ LTR5Hs ZNF729<sup>-/-</sup> (red,  $n = 2$ ) hnPSCs. RNA values are normalized relative to RPL13A. Data represents mean values  $\pm$  SD.



**Extended Data Fig. 8 | ZNF729 binds transposable elements in hnPSCs.**

**a.** *ZNF729* is the most dysregulated KZFP in the blastoids epiblast upon LTR5Hs-repression. Scatter plot depicting expression changes in KZFPs (curated in Krabopedia<sup>99</sup>) in the blastoids epiblast upon LTR5Hs repression (x-axis) vs their expression in hnPSCs (y-axis). FDR 5%. **b.** Histogram representing overrepresented classes of transposable elements (TEs) in the ZNF729-FH bound DNA regions. Only the top 12 families are represented. Y-axis displays

the number of bound TEs from each family. **c.** IGV genome browser capture of ZNF729-FH ChIP-seq signal (top two tracks, purple) or TRIM28 ChIP-seq signal (bottom two tracks, turquoise) in DMSO treated (control) and 24 h dTAG<sup>-1</sup> treated ZNF729-FH at loci containing TEs bound by ZNF729 (SVA\_C, left panel, hg38 chr3:41,980,857-41,986,113; LTR12C, right panel hg38 chr2:38,395,456-38,401,019).



**Extended Data Fig. 9 | ZNF729 exerts activator or repressor function at target genes.** **a.** Low impact of ZNF729-FH depletion on the expression of TEs. MA plot represents TEtranscript<sup>87</sup>-obtained data quantified with DESeq2<sup>83</sup>. Top ZNF729-bound TE families are labeled. **b.** Expression from gene promoters bound or unbound by ZNF729. Box plots show basal hnPSC expression (in TPMs) of genes bound or unbound by ZNF729,  $n = 4$  biological replicates. Boxes show the median and the quartile 1 (Q1) to quartile 3 (Q3) range (interquartile range, IQR). Whiskers minimum corresponds to  $Q1 - 1.5 \times IQR$ , and the maximum to  $Q3 + 1.5 \times IQR$ .  $N = 461$  bound promoters and  $n = 13883$  unbound promoters. Two-sided Wilcoxon test,  $P < 2.2e-16$ . **c.** Stacked bar plot represents number of down- or upregulated genes upon ZNF729-FH depletion for 24 h with dTAG<sup>-1</sup>.

and the percentage of them directly bound by ZNF729 at their promoters ( $-1\text{kb}/+200\text{ bp}$  around the TSS). **d.** Metagenic plots showing TRIM28 ChIP-seq signal over promoters activated or repressed by ZNF729 and over a negative control (TRIM28 bound regions not overlapping with ZNF729). Red represents ZNF729FH DMSO (control) and green dTAG<sup>-1</sup> treated. **e.** PCA analysis of bulk transcriptomes from blastoids, dark spheres formed upon high LTR5Hs repression, and DMSO-treated or dTAG<sup>-1</sup> treated ZNF729-FH hnPSCs. Red shade represents “normal / high” levels of ZNF729 (blastoids, and DMSO treated ZNF729-FH cells). Yellow shade represents “low” ZNF729 levels (dark spheres with high LTR5Hs repression, and dTAG<sup>-1</sup> treated ZNF729-FH hnPSCs). Note that high ZNF729 and low ZNF729 samples separate along PC3.



Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Flow cytometry data were collected on a SONY MA900. qPCR data on a LightCycler 480 II instrument. CUT&RUN and bulk RNA-seq data on a Novaseq 6000 Illumina sequencer. ChIP-seq and single-cell RNA-seq data on a Novaseq X Plus Illumina sequencer. Immunofluorescence images were taken on an Inverted Zeiss LSM 780 confocal microscope. Bright field images on an EVOS FL Imaging System. Blastoids video was taken on a Nikon Eclipse Ti-E. Western blots were revealed on an Amersham ImageQuant 800.
Data analysis	Flow cytometry was analyzed with FlowJo v.10.10.0. Fiji/ImageJ v1.54 was used for microscopy image analysis and measurements. In sequencing data analyses, we used Cutadapt v4.6 to trim Illumina adaptors (in RNA-seq adaptor trimming was performed using Skewer v0.2.2). Bowtie v2.5.1 and Hisat2 v2.2.1 were used to map sequencing data. Transcript alignment was performed with Salmon v1.4.0. Samtools v1.10 to remove PCR duplicates. Peaks were called with MACS3 v3.0.1 Deeptools v3.5.3 to generate bigwigs. Browser captures were obtained in IGV v2.16.0. Differential gene expression analysis was performed in RStudio Build 421 using DESeq2 v1.42.0. ZNF729 transcript in Rhesus was searched using Blast v2.15.0 and Trinity v2.15.2. TETranscripts v2.2.3. Count matrices from PIP-seq data were obtained with Pipeseeker v3.0.5 and CellBender v0.3.0. Downstream analyses were performed in RStudio Build 421 using Seurat v4.4.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Datasets generated in this manuscript can be found at Gene Expression Omnibus (GEO). Accessions: GSE262191, GSE296554, GSE296555, and GSE262329. Datasets were aligned to the human genome assembly hg38. Publicly available datasets analyzed: Petropoulos et al. Human embryo scRNA-seq (ArrayExpress: E-MTAB-3929), Kagawa et al. Blastoids scRNA-seq (GEO GSE177689). Human naïve pluripotent stem cells bulk RNA-seq: naïve\_H9 (GEO GSE131747 and GSE144994), HNES1 and cRH9 GSE123055. Chimpanzee naïve pluripotent stem cells (GEO: GSE264735, genome assembly panTro6 Clint\_PTRv2). Macaque naïve pluripotent stem cells (GEO: GSE61420, genome assembly Mmul\_10 (RheMac10)). Reference human embryo datasets for projections (ArrayExpress E-MTAB-3929, GEO GSE36552, GEO GSE136447, ArrayExpress E-MTAB-9388, ENA PRJEB30442, GEO GSE171820). KRAB zinc finger protein compendium KRABopedia: <https://krabopedia.org/>. RepeatMasker (RRID:SCR\_012954). Isoform resolved human embryo transcriptome (<https://denis-torre.github.io/embryo-transcriptome/#genome-browser>).

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	This work was performed following the 2021 ISSCR Guidelines <sup>50</sup> . The use of blood-derived induced naïve pluripotent stem cells for the experiments described in this manuscript was approved by the Stanford Stem Cell Research Oversight committee (SCRO Protocol number 900).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The sample sizes presented in this manuscript were selected aiming to capture the variability of the experiments based on our laboratory experience. For new approaches, such as measuring blastoid formation potential vs. LTR5Hs activity, for which experimental variability was a concern, we derived tenths of clones instead of the usual 2 or 3 clones per condition, and performed more than two replicates with each clone. For RNA-seq (bulk or single-cell) we collected a sample size sufficient to perform differential gene expression analysis, $n \geq 2$ biological replicates from each sample. For imaging experiments we followed common practices in the developmental biology field (i.e. staining multiple structures from different biological replicates) to capture variability across cell passages and across experiments. For ChIP-seq experiments we performed two independent biological replicates. We deemed this sufficient because the replicates correlated and we had dTAG controls for the specificity.
Data exclusions	In the PIP-seq data, low quality cells were discarded and the specific cut-offs have been included in the Supplementary Table 3.
Replication	All the experiments were biologically replicated and when applicable, at least two clones were used per condition.
Randomization	Blastoids/dark spheres picked for stainings were randomly selected from the pool of blastoids in the plate.
Blinding	We did not apply any blinding approach to our experiments because all samples were treated with cumate. In the case of dTAG experiments blinding is not possible because cells start growing slower.

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

### Antibodies used

#### Primary antibodies:

H3K9me3, CUT&RUN, Abcam ab8898, lot 1063770-1, 0.5 ug  
 HERVK-envelope, Immunostaining, Austral Biological 1811-5, clone 6B7/14, lot MA170221HK, 1:1000  
 KLF17, Immunostaining, Sigma HPA024629, lot 000044816, 1:200  
 GATA3, Immunostaining, R&D Systems AF2605, lot UZQ0223041, 1:200  
 GATA4, Immunostaining, eBioEvan 14-9980-82, clone Ebioevan, lot 2448555, 1:100  
 SUSD2, Immunostaining, R&D Systems MAB90562, clone 1279A, lot CKHB0121031, 1:100  
 Cleaved-CASP3, Immunostaining, Cell Signaling Technology 9661, lot unknown, 1:200  
 NANOG, Immunostaining, Abcam ab109250, clone [EPR2027(2)], lot 1001613-39, 1:100  
 SOX17, Immunostaining, R&D Systems AF1924, lot KGA1021031, 1:200  
 IFI16, Immunostaining, Cell Signaling Technology 14970, clone D8B5T, lot 1, 1:100  
 TROP2 (TACSTD2), Flow cytometry, BD Biosciences 563243, clone 16246, lot 2320729, 1:1000  
 ANPEP, Flow cytometry, BioLegend 301716, clone WM15, lot B395908, 1:200  
 HA-tag monoclonal, Western Blot, Abcam ab18181, clone [HA.C5], 1:1000  
 $\alpha$ -ACTIN, Western Blot, Sigma CBL171-I, clone ASM-1, lot unknown, 0.5ug/ml  
 $\beta$ -ACTIN, Western Blot, Abcam ab49900, clone AC15, lot 1055692-1, 1:5000  
 HA-tag polyclonal ChIP-seq / Western Blot, Abcam 9110, lot GR146572, 5 ug / 1:2000  
 TRIM28, ChIP-seq, GeneTex GTX102226, clone NC32, lot unknown, 5 ug  
 H3K4me3, ChIP-seq, Active motif 39159, lot 22118006, 5 ug  
 H3K27ac, ChIP-seq, Active Motif 39133, lot 31521015, 5 ug

#### Secondary antibodies:

Alexa Fluor Donkey anti-mouse Plus 405, Immunostaining, invitrogen A48257, lot XB342724, 1:500  
 Alexa Fluor Donkey anti-mouse 488, Immunostaining, invitrogen A21202, lot 1796367, 1:500  
 Alexa Fluor Donkey anti-mouse 647, Immunostaining, Life technologies A31571, lot 1757130, 1:500  
 Alexa Fluor Donkey anti-rabbit 405, Immunostaining, invitrogen A48258, lot XA344369, 1:500  
 Alexa Fluor Donkey anti-rabbit 647, Immunostaining, life technologies A31573, lot 1693297, 1:500  
 Alexa Fluor Donkey anti-rat 405, Immunostaining, invitrogen A48268, lot WJ337984, 1:500  
 Alexa Fluor Donkey anti-rat 647, Immunostaining, invitrogen A48272TR, lot XG354854, 1:500

### Validation

All the antibodies have been validated by the vendor for human. Validation websites are indicated:

H3K9me3 <https://www.abcam.com/en-us/products/primary-antibodies/histone-h3-tri-methyl-k9-antibody-chip-grade-ab8898#overlay=images>  
 HERVK-envelope <http://www.australbiologicals.com/index.php?what=catalog&id=338> & PMID:30070637  
 KLF17 <https://www.sigmaaldrich.com/US/en/product/sigma/hpa024629>  
 GATA3 [https://www.rndsystems.com/products/human-gata-3-antibody\\_af2605](https://www.rndsystems.com/products/human-gata-3-antibody_af2605)  
 GATA4 <https://www.thermofisher.com/antibody/product/Gata-4-Antibody-clone-eBioEvan-Monoclonal/14-9980-82>  
 SUSD2 [https://www.rndsystems.com/products/human-susd2-antibody-1279a\\_mab90562](https://www.rndsystems.com/products/human-susd2-antibody-1279a_mab90562)  
 Cleaved-CASP3 <https://www.cellsignal.com/products/primary-antibodies/cleaved-caspase-3-asp175-antibody/9661/applications?index=1&application=IF-IC&type=pdp>  
 NANOG <https://www.abcam.com/en-us/products/primary-antibodies/nanog-antibody-epr20272-ab109250>  
 SOX17 [https://www.rndsystems.com/products/human-sox17-antibody\\_af1924](https://www.rndsystems.com/products/human-sox17-antibody_af1924)  
 IFI16 <https://www.cellsignal.com/products/primary-antibodies/ifi16-d8b5t-rabbit-mab/14970>  
 TROP2 (TACSTD2) [https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv421-mouse-anti-human-trop-2.563243?tab=product\\_details](https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/bv421-mouse-anti-human-trop-2.563243?tab=product_details)  
 ANPEP <https://www.biolegend.com/ja-jp/products/brilliant-violet-421-anti-human-cd13-antibody-8895>  
 HA-tag monoclonal <https://www.abcam.com/en-us/products/primary-antibodies/ha-tag-antibody-hac5-ab18181>  
 $\alpha$ -ACTIN <https://www.sigmaaldrich.com/US/en/product/mm/cbl171i>  
 $\beta$ -ACTIN <https://www.abcam.com/en-us/products/primary-antibodies/hrp-beta-actin-antibody-ac-15-loading-control-ab49900>

HA-tag polyclonal <https://www.abcam.com/en-us/products/primary-antibodies/ha-tag-antibody-chip-grade-ab91110#overlay=images>  
 TRIM28 <https://www.genetex.com/Product/Detail/KAP1-antibody-N3C2-Internal/GTX102226>  
 H3K4me3 <https://www.activemotif.com/catalog/details/39159/histone-h3-trimethyl-lys4-antibody-pab>  
 H3K27ac <https://www.activemotif.com/catalog/details/39133/histone-h3-acetyl-lys27-antibody-pab>

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Peripheral blood cells were donated from consenting volunteers to establish hnPSCs at the University of Tokyo. Official name: PB004 Human iPSC line. Cells were generated in this study: PMID: 26023098
Authentication	STR analysis
Mycoplasma contamination	Cells were routinely tested for mycoplasma. and tested negative
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	No common misidentified cell lines were used in this study.

## Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>

## ChIP-seq

### Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- ☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links <i>May remain private before publication.</i>	GSE296555. Called peaks can be found in Supplementary Table 6, hg38.
Files in database submission	<p>Raw data files</p> <p>INPUT_DMSO_REP1_1.fq.gz INPUT_DMSO_REP1_2.fq.gz            ZNF729-HA_DMSO_REP1_1.fq.gz ZNF729-HA_DMSO_REP1_2.fq.gz            INPUT_dTAG_REP1_1.fq.gz INPUT_dTAG_REP1_2.fq.gz            ZNF729-HA_dTAG_REP1_1.fq.gz ZNF729-HA_dTAG_REP1_2.fq.gz            INPUT_DMSO_REP2_1.fq.gz INPUT_DMSO_REP2_2.fq.gz            ZNF729-HA_DMSO_REP2_1.fq.gz ZNF729-HA_DMSO_REP2_2.fq.gz            INPUT_dTAG_REP2_1.fq.gz INPUT_dTAG_REP2_2.fq.gz            ZNF729-HA_dTAG_REP2_1.fq.gz ZNF729-HA_dTAG_REP2_2.fq.gz            INPUT_DMSO_ZNF729FH_REP1_1.fq.gz INPUT_DMSO_ZNF729FH_REP1_2.fq.gz            TRIM28_DMSO_ZNF729FH_REP1_1.fq.gz TRIM28_DMSO_ZNF729FH_REP1_2.fq.gz            INPUT_dTAG_ZNF729FH_REP1_1.fq.gz INPUT_dTAG_ZNF729FH_REP1_2.fq.gz            TRIM28_dTAG_ZNF729FH_REP1_1.fq.gz TRIM28_dTAG_ZNF729FH_REP1_2.fq.gz            INPUT_DMSO_ZNF729FH_REP2_1.fq.gz INPUT_DMSO_ZNF729FH_REP2_2.fq.gz            TRIM28_DMSO_ZNF729FH_REP2_1.fq.gz TRIM28_DMSO_ZNF729FH_REP2_2.fq.gz            INPUT_dTAG_ZNF729FH_REP2_1.fq.gz INPUT_dTAG_ZNF729FH_REP2_2.fq.gz            TRIM28_dTAG_ZNF729FH_REP2_1.fq.gz TRIM28_dTAG_ZNF729FH_REP2_2.fq.gz            H3K4me3_REP1_1.fq.gz H3K4me3_REP1_2.fq.gz            H3K4me3_REP2_1.fq.gz H3K4me3_REP2_2.fq.gz            H3K27ac_REP1_1.fq.gz H3K27ac_REP1_2.fq.gz            H3K27ac_REP2_1.fq.gz H3K27ac_REP2_2.fq.gz</p> <p>Processed data files</p> <p>INPUT_DMSO_REP1.bw            ZNF729-HA_DMSO_REP1.bw            INPUT_dTAG_REP1.bw            ZNF729-HA_dTAG_REP1.bw            INPUT_DMSO_REP2.bw</p>

ZNF729-HA\_DMSO\_REP2.bw  
 INPUT\_dTAG\_REP2.bw  
 ZNF729-HA\_dTAG\_REP2.bw  
 INPUT\_DMSO\_ZNF729FH\_REP1.bw  
 TRIM28\_DMSO\_ZNF729FH\_REP1.bw  
 INPUT\_dTAG\_ZNF729FH\_REP1.bw  
 TRIM28\_dTAG\_ZNF729FH\_REP1.bw  
 INPUT\_DMSO\_ZNF729FH\_REP2.bw  
 TRIM28\_DMSO\_ZNF729FH\_REP2.bw  
 INPUT\_dTAG\_ZNF729FH\_REP2.bw  
 TRIM28\_dTAG\_ZNF729FH\_REP2.bw  
 H3K4me3\_REP1.bw  
 H3K4me3\_REP2.bw  
 H3K27ac\_REP1.bw  
 H3K27ac\_REP2.bw

Genome browser session  
 (e.g. [UCSC](#))

No longer applicable

## Methodology

Replicates	Two biologically independent replicates per ChIP
Sequencing depth	20-40 M reads
Antibodies	HA tag: Abcam 9110, TRIM28 GeneTex GTX102226, H3K4me3 Active motif 39159, H3K27ac Active Motif 39133
Peak calling parameters	MACS3 software, -q 0.05. Narrow peaks
Data quality	Data quality was assessed by Novogene and by ourselves
Software	Our quality control analysis was performed with FASTP

## Flow Cytometry

### Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation	After 3 days of trophectoderm or hypoblast differentiation cells were used for staining. Cells were pelleted and resuspended in 100 uL of N2B27 supplemented with 10 uM Y-27632
Instrument	SONY MA900
Software	FlowJo v.10.10.0.
Cell population abundance	Percentage of differentiated cells was defined using the negative control.
Gating strategy	Gating example can be found in Extended Data Figure 4E and 4H. Debris and cellular aggregates were removed from analysis via FSC-A/SSC-A gating following by FSC-A/FSC-W and SSC-A/SSC-W gating. Positive signal was defined as cells not overlapping with the negative control (unstained cells).

- ☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.